

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:36:43 ; Search time 53 seconds
(without alignments)
2009.195 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNWLVLVLLCGKHH.....PRVKRPQLMQVHPKPTPSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 383519 seqs, 101223694 residues

Total number of hits satisfying chosen parameters: 383519

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5617	100.0	1052	10	US-09-060-854B-7
2	5617	100.0	1052	10	US-09-891-711-4
3	5487	97.7	1052	10	US-09-891-711-6
4	337.5	6.0	357	9	US-09-837-235-15
5	337.5	6.0	379	9	US-09-813-408-11
6	336.5	6.0	1079	9	US-09-112-488-39
7	335.5	6.0	379	10	US-09-920-118-14
8	332.5	5.9	381	10	US-09-920-118-16
9	329.5	5.9	379	9	US-09-813-408-10
10	325	5.8	380	9	US-09-813-408-16
11	324.5	5.8	379	9	US-09-813-408-13
12	323.5	5.8	275	1	US-08-322-678-8
13	323.5	5.8	275	9	US-10-033-325-2
14	323.5	5.8	275	10	US-09-060-854B-4
15	322.5	5.7	378	9	US-09-813-408-14
16	322	5.7	382	9	US-09-813-408-7
17	322	5.7	382	9	US-10-033-325-2
18	320.5	5.7	275	9	US-09-813-408-18
19	319.5	5.7	372	9	US-09-813-408-9

ALIGNMENTS

RESULT 1

US-09-060-854B-7
; Sequence 7, Application US/09060854B
; Patent No. US20020081703A1

; GENERAL INFORMATION:

; APPLICANT: Estell, David Aaron

; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical

; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002

; TITLE OF INVENTION: Proteins

; FILE REFERENCE: GC532

; CURRENT APPLICATION NUMBER: US/09/060,854B

; CURRENT FILING DATE: 1998-04-15

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 7

; LENGTH: 1052

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-060-854B-7

Query Match	100.0%;	Score	5617;	DB	10;	Length	1052;
Best Local Similarity	100.0%;	Pred. No.	0;				
Matches	1052;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MKLVNWLVLVLLCGKHHDRLEKKSPEKAPCGCSHLTKVFSSVWEYIVAF	60				
Db	1	MKLVNWLVLVLLCGKHHDRLEKKSPEKAPCGCSHLTKVFSSVWEYIVAF	60				
Qy	61	NGYFTAKARNSFISSALKSSEVDNWRIPRNNSSDYPDFEVIQIKEKQAGLLTLEH	120				
Db	61	NGYFTAKARNSFISSALKSSEVDNWRIPRNNSSDYPDFEVIQIKEKQAGLLTLEH	120				
Qy	121	PNIKRVTPQKRVFRSLKYAESDTPVPCNETRWQKWSRPLRRASLSLGSFGFWHATGRH	180				
Db	121	PNIKRVTPQKRVFRSLKYAESDTPVPCNETRWQKWSRPLRRASLSLGSFGFWHATGRH	180				
Qy	181	SSRRLRAIPRQVAQTLDVLMQMGYTGANVRVAVDTGLSEKHPHFKNVKERTNWTNE	240				
Db	181	SSRRLRAIPRQVAQTLDVLMQMGYTGANVRVAVDTGLSEKHPHFKNVKERTNWTNE	240				
Qy	241	RTLDDGLGHGTFVAGVVIASMRQCQGFAPDAELHIFRVTNNQVSYTSWFLDAENYALKK	300				
Db	241	RTLDDGLGHGTFVAGVVIASMRQCQGFAPDAELHIFRVTNNQVSYTSWFLDAENYALKK	300				

Db	241	RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRVFTNNQVSYWFLDAFNAILKK	300
Qy	301	IDVLNLSIGGPDFMDHFFVDKQWELTANNVIMVSAIGNDGPLYGLTLNPNADQMDVIGVG	360
Db	301	IDVLNLSIGGPDFMDHFFVDKQWELTANNVIMVSAIGNDGPLYGLTLNPNADQMDVIGVG	360
Qy	361	IDFEDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPV	420
Db	361	IDFEDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPV	420
Qy	421	AGAVTLVSTVQKRELVPASMKQALIASARLPQVNMFEQGHKGLDLILRAYOILNSYKP	480
Db	421	AGAVTLVSTVQKRELVPASMKQALIASARLPQVNMFEQGHKGLDLILRAYOILNSYKP	480
Qy	481	QASLSPSYIDLTECPYMWPCYCSQPIYYGGMTVVNTILNGMGVTRGVDPQWQYLPQ	540
Db	481	QASLSPSYIDLTECPYMWPCYCSQPIYYGGMTVVNTILNGMGVTRGVDPQWQYLPQ	540
Qy	541	NGDNIEAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE	600
Db	541	NGDNIEAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE	600
Qy	601	QTSVKLP IKVKIIPTPPSRKVLMDQYHNLRYPPGYPRNLRMKNDPLDWNQDHIHTN	660
Db	601	QTSVKLP IKVKIIPTPPSRKVLMDQYHNLRYPPGYPRNLRMKNDPLDWNQDHIHTN	660
Qy	661	FRDMTOHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEELAKLRDNDGLSL	720
Db	661	FRDMTOHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEELAKLRDNDGLSL	720
Qy	721	VIFSDWNTSVMRKVKFYDENTROMMPDTGGANIPALNELLSSVNMGFSGLYEGEFTL	780
Db	721	VIFSDWNTSVMRKVKFYDENTROMMPDTGGANIPALNELLSSVNMGFSGLYEGEFTL	780
Qy	781	ANHDMYASGCSIAKPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGBGR	840
Db	781	ANHDMYASGCSIAKPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGBGR	840
Qy	841	IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM	900
Db	841	IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM	900
Qy	901	EGNHLHRYSKVLEAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQV	960
Db	901	EGNHLHRYSKVLEAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQV	960
Qy	961	VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ	1020
Db	961	VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ	1020
Qy	1021	INKAKSRPKRRKPRVKRQPMQOVHPKTPSV	1052
Db	1021	INKAKSRPKRRKPRVKRQPMQOVHPKTPSV	1052

RESULT 2
 ; Sequence 4, Application US/09891711
 ; Patent No. US20020082404A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Pharmacia AB
 ; TITLE OF INVENTION: Promoter Sequences
 ; FILE REFERENCE: 00130
 ; CURRENT APPLICATION NUMBER: US/09/891,711
 ; CURRENT FILING DATE: 2001-06-26
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 4
 ; LENGTH: 1052
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 us-09-891-711-4

Query Match	100.0%;	Score 5617;	DB 10;	Length 1052;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1052;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKLVNILLVLLVLCGKKGHLDRLEKKSFEKAPCPGCSHLTLKVBFSSVTVVEYIVAF	£0	
Db	1	MKLVNILLVLLVLCGKKGHLDRLEKKSFEKAPCPGCSHLTLKVBFSSVTVVEYIVAF	£0	
Qy	61	NGYFTAKARNSFTSSALKSEVDNRRIIPRNPSSDYPSPDFVIOIKKQKAGLLTLDH	120	
Db	61	NGYFTAKARNSFTSSALKSEVDNRRIIPRNPSSDYPSPDFVIOIKKQKAGLLTLDH	120	
Qy	121	PHIKVTPQKRVPRSLKYAESDPTVPCNETRMSQKWOSSRPLRRASLSLGSQFHHAYGRH	180	
Db	121	PHIKVTPQKRVPRSLKYAESDPTVPCNETRMSQKWOSSRPLRRASLSLGSQFHHAYGRH	180	
Qy	181	SSRRLLRAIPROVAQTLQADVLQMGYTGANVRVAVEDTGLSEKHPHFKNVKERTNMTNE	240	
Db	181	SSRRLLRAIPROVAQTLQADVLQMGYTGANVRVAVEDTGLSEKHPHFKNVKERTNMTNE	240	
Qy	241	RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRVFTNNQVSYWFLDAFNAILKK	£00	
Db	241	RTLDGLGHGTFVAGVIASRECOGFAPDAELHIFRVFTNNQVSYWFLDAFNAILKK	£00	
Qy	301	IDVLNLSIGGPDFMDHFFVDKQWELTANNVIMVSAIGNDGPLYGLTLNPNADQMDVIGVG	£60	
Db	301	IDVLNLSIGGPDFMDHFFVDKQWELTANNVIMVSAIGNDGPLYGLTLNPNADQMDVIGVG	£60	
Qy	361	IDFEDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPV	420	
Db	361	IDFEDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPV	420	
Qy	421	AGAVTLVSTVQKRELVPASMKQALIASARLPQVNMFEQGHKGLDLILRAYOILNSYKP	480	
Db	421	AGAVTLVSTVQKRELVPASMKQALIASARLPQVNMFEQGHKGLDLILRAYOILNSYKP	480	
Qy	481	QASLSPSYIDLTECPYMWPCYCSQPIYYGGMTVVNTILNGMGVTRGVDPQWQYLPQ	£40	
Db	481	QASLSPSYIDLTECPYMWPCYCSQPIYYGGMTVVNTILNGMGVTRGVDPQWQYLPQ	£40	
Qy	541	NGDNIEAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE	£00	
Db	541	NGDNIEAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE	£00	
Qy	601	QTSVKLP IKVKIIPTPPSRKVLMDQYHNLRYPPGYPRDLNRMKNDPLDWNQDHIHTN	£60	
Db	601	QTSVKLP IKVKIIPTPPSRKVLMDQYHNLRYPPGYPRDLNRMKNDPLDWNQDHIHTN	£60	
Qy	661	FRDMTOHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEELAKLRDNDGLSL	£20	
Db	661	FRDMTOHLRSMGYFVEVLGAPTCFDSQYGTLLMVDSEEEYFPEELAKLRDNDGLSL	£20	
Qy	721	VIFSDWNTSVMRKVKFYDENTROMMPDTGGANIPALNELLSSVNMGFSGLYEGEFTL	£80	
Db	721	VIFSDWNTSVMRKVKFYDENTROMMPDTGGANIPALNELLSSVNMGFSGLYEGEFTL	£80	
Qy	781	ANHDMYASGCSIAKPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGBGR	£40	
Db	781	ANHDMYASGCSIAKPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGBGR	£40	
Qy	841	IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM	£00	
Db	841	IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM	£00	
Qy	901	EGNHLHRYSKVLEAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQV	£60	
Db	901	EGNHLHRYSKVLEAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQV	£60	
Qy	961	VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ	£020	
Db	961	VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVYLAFFVQ	£020	
Qy	1021	INKAKSRPKRRKPRVKRQPMQOVHPKTPSV	1052	

Db 182 GSGQSWIINGIEMAISSNNNOVINMSLGGTGTALKTVVDKA---VSSGIIVAAAGNE 238

QY 340 GP--LYGTLPNADQMDVIGVGGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYGAG 397

Db 239 GSGGSTVGYPAKYESTIAGVANSNQRASFSSAGS---EL-----DVMAPGVS 286

QY 398 VRGSGVGGCRALSGTSVASPVVAGAVTLVSTVQKRELVPNPMKQALIASARLPVGN 457

Db 287 IGSTLPGGTYGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 341

QY 458 MFEQGHGKLDLLRAYO 473

Db 342 SFYCKGLINVOAAQ 357

RESULT 5

US-09-813-408-11

; Sequence 11, Application US/09813408

; Publication NO. US20030049619A1

; GENERAL INFORMATION:

; APPLICANT: Delagrave, Simon

; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides

; FILE REFERENCE: HER0041

; CURRENT APPLICATION NUMBER: US/09/813,408

; CURRENT FILING DATE: 2001-03-21

; NUMBER OF SEQ ID NOS: 85

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 11

; LENGTH: 379

; TYPE: PRT

; ORGANISM: Bacillus licheniformis

US-09-813-408-11

Query Match 6.0%; Score 337.5; DB 9; Length 379;

Best Local Similarity 25.7%; Pred. No. 1.9e-20;

Matches 126; Conservative 66; Mismatches 169; Indels 129; Gaps 17;

QY 1 MKLVNILLVLLCGKKHLGRLEKKSFEKAPCGCSHLLTKVRESSV-----V 52

Db 2 MRKSEWLCMTALM-----LVFTMAFSDSASAAQAKNV 36

QY 53 EYEYIVAF--NGYFTAKRNSFISSAKASVD--NRIIPNPPSSDYPSEFVIOIKERQ 110

Db 37 EKDIIVGKSGVKTASVKKDIINES--GGKVDKQFRIINAARAKLD-----80

QY 111 KAGLTLEDHPNKRVTPOKVRSLKYAESDPTVPCNETRWSQKSSRPLRRASLSLG 170

Db 81 KEALEEVKNDPDV-----ATVEED-----99

QY 171 SGFWHATGRHSRRLLRAIPROVAQTLOADVLMOMGYTGANYRVAVFDTGLSEKHPHFKN 230

Db 100 -----HYAHALAQTVPYGI--FLIKADKVAQGYKGANVKAVALDTGIQASHPDNAV 149

QY 231 VKERTNWTNERTLDDGLGHGTFTVAGVIASMRQO---GFAPDAELHIFRVFTNNOVSYTS 287

Db 150 VGSASFVAGEAYNTDGNHGTHTVAGTVAALDNTTGLVGNPVNLSYAKVLNLSGSGSYS 209

QY 288 WFLDAFNAILKIDVLNLSIGGPDFMD--HPFVDKRWELTANNVIMVSAIGNDGDL--Y 343

Db 210 GIVSGIEWATTNGMDVINSLGSPSGSTAMQAQVDNAY---ARGVVVVAAGNSGSGNT 266

QY 344 GTLNPADOMDVGIGDIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGVRSV 403

Db 267 NTIGTPAKYDVIAGVADSNRASFSSVCA--EL-----EVMAPGAGYSTTP 314

QY 404 KGCRLSGTSVASPVVAGAVTLVSTVQKRELVPNPMKQALIASARLPVGNMFEQGH 463

Db 315 TSTYATLNGTSMASPHVAGAAALILS---RHPNLSASQVRNRLSLSTATYLG--SSFYCK 369

QY 464 GKLDLLRAYO 473

Db 370 GLINVEAAQ 379

RESULT 6

US-10-112-488-39

; Sequence 39, Application US/10112488

; Publication NO. US20030082746A1

; GENERAL INFORMATION:

; APPLICANT: KIKUCHI, Yoshiaki

; APPLICANT: DATE, Masayo

; APPLICANT: UMEZAWA, Yukiko

; APPLICANT: YOKOYAMA, Keiichi

; APPLICANT: MATSUI, Hiroshi

; TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSLUTAMINASE

; FILE REFERENCE: 219286USOCONT

; CURRENT APPLICATION NUMBER: US/10/112,488

; CURRENT FILING DATE: 2002-04-01

; PRIOR APPLICATION NUMBER: PCT/JP00/06780

; PRIOR FILING DATE: 2000-09-29

; PRIOR APPLICATION NUMBER: JP2000-280098

; PRIOR FILING DATE: 2000-06-28

; PRIOR APPLICATION NUMBER: JP11-280098

; PRIOR FILING DATE: 1999-09-30

; NUMBER OF SEQ ID NOS: 70

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 39

; LENGTH: 1079

; TYPE: PRT

; ORGANISM: Streptomyces albobogiseolus

US-10-112-488-39

Query Match 6.0%; Score 336.5; DB 9; Length 1079;

Best Local Similarity 22.3%; Pred. No. 1.3e-19;

Matches 175; Conservative 113; Mismatches 284; Indels 211; Gaps 36;

QY 162 LRRASLSL-----GSGFWHA-----TGRHSRRLLRAIPROVAQTLOADV---201

Db 124 VRRTLSLNADAVTQPEAGAELEWAVTDGDRTASGVARVWLDGVRKASLDTSVQIGTP 183

QY 202 -LWOMGYTGANYRVAVFDTGLSEKHPHF--NVKERTNWTNERTLDDGLGHGTFTVAGVIA-258

Db 184 KWEAGYDGKGVKIAVLDTGVDATHPDLKGQVTSKNTSAPTTGVDVGVGHVTHASIAAG 343

QY 259 ----SMRECOGFAPDAELHIFRVFTNNOVSYTSWFLDAFNAILKIDVLNLSIGGPDFM 314

Db 244 TGAQSKGYIKGVAFCAKILNGKVLDDAGGDDSGILAGMEWAAQAQADIVNLSLG---M 300

QY 315 DHPFVD---KWELTA--NNVIMVSAIGNDGLYGTLPNADQMDVIGGIDFEDNIAR 369

Db 301 DTPETDPLEAAVDKLSAEKGLFAPAGNEG--QSGSPGSADSALTGVADDKDLAD 358

QY 370 FSSRGMTTWELPGYGRMKPDIVTYGAGVRSVKG-----CCRALSGTSVAS 417

Db 359 FSTGTPRL-----GDGAVKPDLTAPGVDTITAAKAGNDIAKFEVGERPAGTMTIGTSMAT 413

QY 418 PVVAGAVTLVSTVQKRELVPNPMKQALIASARLPVGNMFGHKGKLDLLRAYOILNS 477

Db 414 PHVAGAAALL---KQHPKPEWKAELKAGALTASTKD--GKYTPFQSGRQVQVDKAITQTVI 469

QY 478 YKPOASLSPSYIDTECPYMWPYC-----SOPYYGGMPTVVAVNTILNMGVGTGRIVDKP 532

Db 470 AEP-----VLSFGVQOWPHADKPVTKKLTATYRNLGT--EDVTLKLTATG-----514

QY 533 DMQPYLPQNGDNIEVAF---SYSSVLWPMGSLAISV--TKKAASWEGIAOCHVMTVA 588

Db 515 -----PKGAAPAGFTLGLASTLTVPANGTASVDVTDTRLGGAVDGTYSAYVATGA 567

QY 589 SPA-----EFESKNGAEQTSVKPLPIKVIPIPPBS-----KRVL 624

Db 568 GGSVRTAAAVREVESTN-----VTLKLDLDSGKATANYAYLSGLTGLGKDRS 616

QY 625 WDOYH-----NLRYPPGYPFRD--NLRMKNDPLDNGDHHTNFRDMYQHLRSMGYFVEVL 678


```

; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813.408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 10
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-813-408-10

```

```

Query Match      5.9%; Score 329.5; DB 9; Length 379;
Best Local Similarity 25.3%; Pred. No. 9.4e-20;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

QY 1 MKLVNILLVLLVLCGKHLGDRLEKKSFEKAPCGCSHLTKLVEFSSTV-----V 52
DB 2 MRKKSFWLGLMTAFM-----LVFTMAFSDSASAAQPKNV 36

QY 53 EYEYIVAF-NGYETAKARNSFTSSALKSSSEVD-NWRIIPRNPSSDYPDFEVIQKEQ 110
DB 37 EXDIYVGRKSVGTASVKKDIKES--GGKVDKQFRIINAAKALD-----80

QY 111 KAGLLTLEDHPNTRKVTPOKRVFSLKYAESDPTVPCNETRWSQKWSRPLRRASLSLG 170
DB 81 KEALKEVNDPDV-----AYVEED-----99

QY 171 SGFWHATGRSSRRLLRAIPROVAOTLQADVLMQMGYTCANRVAVFDTGLSEKHPFKN 230
DB 100 -----HYAHLAQTYPGI-PLIKADKVAQOGFGKANVKVAVLDTGIOASHPDNLV 149

QY 231 VKERTNWTNERTLDDGLGHGTFVAGVIASMRQC--GFAPDAELHIFRYFTNNQVSYTS 287
DB 150 VGGASVFAGEAYNTDGNHGTHVAGTVAALDNTTGLVGVAPSVSLYAVKVLNSSGGTYS 209

QY 288 WFLDAFNAILKIDVNLISIGGPDFMD--HPFVDKRWELTANNVIMVSAIGNDGL--Y 343
DB 210 GIVSGIEWATTNGMDVINMISLGGSPGSGTAMQAVDNAY---ARGVVVAAAGNSGSGNT. 266

QY 344 GTLNNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVYAGVRGSGV 403
DB 267 NTIGYPKIDSVIAGVADNSNRASFSSVGA---EL-----EVMAPGAGVYSTYP 314

QY 404 KGGCRALSGTSVSPVAVAGATLLVSTVQKRELNVNPSMKAQALASARLLPGVNMEEQGH 463
DB 315 TSYATLNGTSMASPHVAGAAALILS--KHPNLSAQVRNLSSTATYLG--SSFYTGK 369

QY 464 GKLDLLRAYQ 473
DB 370 GLINVEAAQA 379

```

```

RESULT 10
US-09-813-408-16
; Sequence 16, Application US/09813408
; Publication NO. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813.408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 16
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Bacillus subtilis var. natto
US-09-813-408-16

```

Query Match 5.8%; Score 325; DB 9; Length 380;

```

Best Local Similarity 27.0%; Pred. No. 2.3e-19;
Matches 119; Conservative 66; Mismatches 148; Indels 108; Gaps 20;

QY 48 SSTVVEYEVIAFNGYFTAKARNSFTSSALKSSSEVDNWRIPRNPSSDYPDFEVIQIK 207
DB 33 SST--EKYIVGF-----KQTMASMSAKKKDVI-----59

QY 108 EKOKAGLLTLE-DHPNTRKVTPOKRVFSLKYAESDPTVPCNETRWSQKWSRPLRRAS 166
DB 60 -SEKGGKVKQFKYVNAARATLDEKAVKELK---KDPVA-----95

QY 167 LSLGSGFWHATGRSSRRLLRAIPROVAOTLQADVLMQMGYTCANRVAVFDTGLSEKHP 226
DB 96 -----YVEEDIAHEVAQSPYGISQ-IRALHSQYGTYSNKKVAVIDSGDSSHP 146

QY 227 HFKNKERTNWTNERT--LDDGLGHGTFVAGVIASMR--COGFAPDAELHIFRYFTNN 281
DB 147 DL-NVRGASVFPSENPYQDSSGHGTHVAGTIAALNNSIGVLGVAPASLAVKVLDTST 205

QY 282 QVSYTSWFLDAFNAILKIDVNLISIGGPDFMD--HPFVDKRWELTANNVIMVSAIGND 339
DB 206 GSGQYSLWINGIEWISNNMDVINMISLGGPTGSTALTATVVDKA---VSSGIVVAAAAGNE 262

QY 340 GP--LYGTNNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVYTAG 397
DB 263 GSGSGTSTVGPAPKPTSTTAVGAVNSNRASFSSVGS---EL-----DVMAPGVS 310

QY 398 VRGSGVRGKC-RALSGTSSVSPVAVAGATLLVSTVQKRELNVN---ASMKAQALASARR 452
DB 311 IQSSTPLGTYGATNGTSMATH-VAGAAALILS-----HPTWTNAQVRDLSESTATY 361

QY 453 LFCVNMFEQGHGKLDLLRAYQ 473
DB 362 LG--NSFYGKGLINVQAAQA 380

```

```

RESULT 11
US-09-813-408-13
; Sequence 13, Application US/09813408
; Publication NO. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813.408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 379
; TYPE: PRT
; ORGANISM: Bacillus licheniformis
US-09-813-408-13

Query Match      5.8%; Score 324.5; DB 9; Length 379;
Best Local Similarity 24.8%; Pred. No. 2.5e-19;
Matches 123; Conservative 66; Mismatches 167; Indels 139; Gaps 17;

QY 1 MKLVNILLVLLVLCGKHLGDRLEKKSFEKAPCGCSHLTKLVEFSSTV-----V 52
DB 2 MRKKSFWLGLMTALM-----LVFTMAFSDSASAAQPKNV 36

QY 53 EYEYIVAF-NGYETAKARNSFTSSALKSSSEVD-NWRIIPRNPSSDYPDFEVIQKEQ 110
DB 37 EXDIYVGRKSVGTASVKKDIKES--GGKVDKQFRIINAAGKALD-----80

QY 111 KAGLLTLEDHPNTRKVTPOKRVFSLKYAESDPTVPCNETRWSQKWSRPLRRASLSLG 170
DB 81 KEALKEVNDPDV-----AYVEED-----59

QY 171 SGFWHATGRSSRRLLRAIPROVAOTLQADVLMQMGYTCANRVAVFDTGLSEKHPFKN 230

```

Db 100 -----HVAHVLGQTPYGI-PLIKADKVOAGFKGANVKVAVLDTCIQASHPDNLV 149
QY 231 VKERTNWTNRTLDGGLGCTFFVAGVIASMBRCQ---GFAPDAELHIFRVTNNQVSYTS 287
Db 150 VGASFVAGPAYWTDGNGHGVAGTVAALONTGVLGVAPSVLSYAVKVLNLSGSGSYS 209
QY 288 WFLDAFNAYAILKIDVNLISIGPDF-----MDHPFVDKVMWELTANNVIMVSAIGNDG 340
Db 210 AIVSGLEWATTTGMDVINLSLGCASVSTAMKQAVDHAY-----ARGAVVSSAGNSG 261
QY 341 PL--YGTLLNPADQMDVIGVGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGV 398
Db 262 SSGTNTIGYPAKYDSVIAVGAVDNSNRASFSSVGA---EL-----EYMAPGAGV 309
QY 399 RGSVGGCCRALSGTSSVSPVAVGAVTLLVSTVQKRELNPASMKQALIASARRLPGVNM 458
Db 310 YSTYPTNTATLNGTSMASHVAGAAALILS-----KHPNLSASQVTRLSRTATYLG--SS 364
QY 459 FEQHGKLDLLRAYQ 473
Db 365 FSYGRGLINVEAAQA 379

RESULT 12

US-08-322-678-8
; Sequence 8, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-678-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 1.8e-19;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
QY 187 RAIPROVATQADVLQMGYTGANVRVAVFDGLSEKHPKFKNKERTNWTNERT--LD 244
Db 2 QSVPYGISQ-IPKALHSQGYTGSNKKVAVIDSGIDSSHPDL-NVRGGASFVPSNTPYQ 59

QY 245 DGLGHGTFVAGVIASMBRE---COGFAPDAELHIFRVTNNQVSYTSWFLDAFNAYAILKKI 301
Db 60 DGSSHCHTHVAGTAAALNNSIGVLGVSFASLYAVKVLDDTSGSGQYSWIINGIEWALSNNM 119
QY 302 DVNLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGNDGP--LYGTLLNPADQMDVIG 357
Db 120 DVINMSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNEGSGSSTSTVGYPAKYPSTIA 176
QY 358 VGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVKGCGCRALSGTSVAS 417
Db 177 VGVAVSSNRASFSSAGS---EL-----DVMAPGVSIQSTLPGGTGYGAYNGTSMAT 224
QY 418 PVVAGAVTLLVSTVQKRELNPASMKQALIASARRLPGVNMFEQHGKLDLLRAYQ 473
Db 225 PVVAGAAALILS---KHPWTNAQVRDRLESTATYLG--NSFYFGKGLINVOVAAAQ 275

RESULT 13

US-10-033-325-4
; Sequence 4, Application US/10033325
; Publication No. US20030073222A1
; GENERAL INFORMATION:
; APPLICANT: Poulouse, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nadherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baeck, Andre C.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2-C1
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; PRIOR FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-033-325-4

Query Match 5.8%; Score 323.5; DB 9; Length 275;
Best Local Similarity 31.1%; Pred. No. 1.8e-19;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
QY 187 RAIPROVATQADVLQMGYTGANVRVAVFDGLSEKHPKFKNKERTNWTNERT--LD 244
Db 2 QSVPYGISQ-IPKALHSQGYTGSNKKVAVIDSGIDSSHPDL-NVRGGASFVPSNTPYQ 59
QY 245 DGLGHGTFVAGVIASMBRE---COGFAPDAELHIFRVTNNQVSYTSWFLDAFNAYAILKKI 301
Db 60 DGSSHCHTHVAGTAAALNNSIGVLGVSFASLYAVKVLDDTSGSGQYSWIINGIEWALSNNM 119
QY 302 DVNLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGNDGP--LYGTLLNPADQMDVIG 357
Db 120 DVINMSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNEGSGSSTSTVGYPAKYPSTIA 176
QY 358 VGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRSVKGCGCRALSGTSVAS 417
Db 177 VGVAVSSNRASFSSAGS---EL-----DVMAPGVSIQSTLPGGTGYGAYNGTSMAT 224
QY 418 PVVAGAVTLLVSTVQKRELNPASMKQALIASARRLPGVNMFEQHGKLDLLRAYQ 473

Qy	258	ASMBRE---	CCGFADPAE	LHIERFVT	NNOVST	SFWLDA	FNAYAIL	KKIDYLN	LSIGSPDM	314
			:	:	:	:	:	:	:	
Db	184	AALNNSIG	VGVADPNAE	LYAVKVL	CANGSG	VSSIAOGL	QWTAQNN	IHVANLS	GS----	239
			:	:	:	:	:	:	:	
Qy	315	DHPFVDK	WELTAN---	NVMYS	AI	GNDCG	PLGYCT	LNNPAD	MDM	369
			:	:	:	:	:	:	:	
Db	240	--PVGSOT	LELAVNAQ	NAGVLL	VAAATG	NGSG--	GTVSYP	ARYANAL	AVGATD	295
			:	:	:	:	:	:	:	
Qy	370	FSSRGMT	TWELPGGY	GRMKP	DI	VTVY	GAGVRS	GVKGC	RALSGT	429
			:	:	:	:	:	:	:	
Db	296	FSQYGTGL	-----	NIVAPG	VIQSTY	PGNRYAS	LSGTSM	ATPHV	AGVAALV---	341
			:	:	:	:	:	:	:	
Qy	430	TVQKREL	VNPASMK	QALIAS	ARRLP	GVNMF	EQG	462		
			:	:	:	:	:	:	:	
Db	342	--KOKNP	SWSNFQIR	ROHLT	ST	STATS	IGNSNO	FQSG	373	

```

RESULT 17
US-10-033-325-2
; Sequence 2, Application US/10033325
; Publication No. US20030073222A1
; GENERAL INFORMATION:
; APPLICANT: Poulos, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nacherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baeck, Andre C.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: G5502-2-C1
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; PRIOR FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Bacillus amyloliquefaciens
US-10-033-325-2

```

Query Match	5.7%	Score 322;	DB 9;	Length 382;														
Best Local Similarity	26.2%;	Pred. No. 4.2e-19;																
Matches 119;	Conservative 57;	Mismatches 142;	Indels 136;	Gaps 18;														
QY	1	MLKLVNWL	LLVLLCGK	KHGLDRLEKK	SFEKAPCPC	SHLT	LVKEFS	TVV-----	52									
Db	1	MRGKKV	MSILFALA	-----	LIITMA	FSGSS	QAQAKSN	35										
QY	53	-EYEV	IVARNGY	FTAKARNS	FISALK	SSVDNW	RIIPRN	NPSSDP	SDFEVIQ	KEKQ 111								
Db	36	GEKKY	IVGF-----	KQT	MTS	MAAKK	DYI-----	SEK	63									
QY	112	AGLLT	LE-DHP	NIKRTVP	QKRV	PSRLKYA	ESDTPV	PCNETR	WQKWSR	PLRRAS	LSLG 170							
Db	64	GGKVQ	KQFY	XVDAA	SVTLNEK	AVKELK-----	KDPSVA-----	96										
QY	171	SGFW	HATG	RHSSR	LLRLRA	IPROVA	QTLQ	ADLVLMQ	MGYTG	ANRV	VAVFDT	GLSEKH	PHFK- 229					
Db	97	-----	YVEED	HVAHAY	AQSPY	GYSQ-IP	KAL	HSQGYT	GSN	KVAV	IDSGIS	DHDLK	V 151					
QY	230	-----	NV	KERT	NTN	ERT	LD	DGLGH	GTF	VAG	VI	ASMR-----	COG	FAP	DAEL	HI	FR	VTNN 281

```

152 ASGASMPVSETN-----PFQDNNSHGHTVAGTVAALNNSIGVLGVAPSASLYAVKVLGAD 206
282 QVSYTSWFLDAFNAYLLKKIDVLNLSIGGPDFMD--HPFVDKQWELTANNVIMVSAIGND 339
207 GSGOYSWIIINGIEWAIANNMDVINNSLGGPSSGAALKAAVDKA---VASGVVVVAAAGNE 263
340 GP---LYGTLLNPADQMDDIVYGGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
264 GTSGSSTVGYPGKYPSVIAVCAVDSSNQRAFSSVGP---EL-----DVMAPGVS 311
398 VRGS--GVKGCRRALSGTSVAPSVVAGAVTLLVS 429
312 IQSTLPGNKYK--AYNGTSMASPRVAGAAAILIS 343

```

```

RESULT 18
US-09-813-408-18
; Sequence 18, Application US/09813408
; Publication No. US2003004961A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Marrs, Barry
; TITLE OF INVENTION: Methods For The Synthesis
; OF Polynucleotides
; FILE REFERENCE: HPR0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus pumilus
US-09-813-408-18

```

Query Match	5.7%	Score	320.5	DB	9	Length	275
Best Local Similarity	31.1%	Pred. No.	3.3e-19				
Matches	92	Conservative	55	Mismatches	116	Indels	31
Gaps							
QY	187	RAIPROVAQTLQADVLQMGGYTGANRVAVFDTGLSEKHPHFKNVKERTNWTNERT--LD	244				
Db	2	QSPVGGISQ-IKAPALHSQGYTGNNKVAVIDSGIDSSHDPDL-NVRGGASFVPSETNPQY	59				
QY	245	DGLGHGTFVAGVIASMRE--CQGFADPAELHIFRVFTNNQVSYTSWFLDAFNAYAILKKI	301				
Db	60	DGSSHGHVTHAGTIAALNNSGVLGVAFASASLYAVKVLDTSGSGQGYSWIINGIEWAISNMN	119				
QY	302	DVLNLSTGGPDFMD--HPFVDKWELTANNYIMVSAIGNDGP--LYGLTNLPADQMDFIG	357				
Db	120	DVTNMSLGGPTGSTALKTVVDKA--VSSGIVWAAAGNEGSSGSTSTGVGPAKYPTSTIA	176				
QY	358	VGGIDFDNIDTARFSSRGMTTWELPGGYGRMKPDIVTYGAGVYRGSGVKGCRLASTSVAS	417				
Db	177	VGAVNSANQRASFSAGS--EL-----DVMAFCVSIQSTLPGGTYGAYNGTSMAT	224				
QY	418	PVVAGAVTLLVSTVQKRELNPASMKQALIASARLLPGVNNMFEQHGKLDLLRAYQ	473				
Db	225	PHVGAALILS---KHEPTWTAQVORLESTATILG--SSFYTGKGLINQVQAAQ	275				

```

RESULT 19
US-09-813-408-9
; Sequence 9, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85

```


; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 9

; LENGTH: 372

; TYPE: PRT

; ORGANISM: Bacillus licheniformis

US-09-813-408-9

Query Match 5.7%; Score 319.5; DB 9; Length 372;
Best Local Similarity 24.6%; Pred. No. 6.6e-19;
Matches 119; Conservative 66; Mismatches 170; Indels 129; Gaps 16;

QY 1 MKLVNMLLLVLLVLLGCKHLLGDRLEKSKPEKAPCPGCSHLTLKVEFSSTV-----V 52

DB 2 MRKSEFLGMLTAFM-----LVETMAFSDSASAAQPAKNV 36

QY 53 EYEXIVAF-NGYFTAKARNSFISSALKSSEVDNWRRIIPRNPSSDYPSPDEVIQIKQK 111

DB 37 EKDYLVGFKSVKTSVKKDIKES--GGKVD-----KQPRIIN----- 73

QY 112 AGLLTLEDHPNIRKVTPOKVFERSIKYAESDPVPCNETRMSQKWSSRLRSLSLGS 171

DB 74 -----AAKAKIDEALKVKNDPVAYVED----- 97

QY 172 GWHTATGHSRRLRAIPROVAOTLQADVLWQGYTGANTRVAVFTGLSEKHPHFKNV 231

DB 98 -----HVAHALAQTVPIGL-PLIRADKVOAQGFKANVAVLDTGQASHDLNVV 148

QY 232 KERTNWTNERTLDGLGHGTFVAGVIASMRQ---GFAPDAELHIFRFTNNQVSYTSW 288

DB 149 GGASEVAGEAYNTDNGHGHTRVAGTVAALDNTTGVLGAPSVSLYAVKVLNLSGSGSYSG 208

QY 289 FIDAFNVAILKKIDVLNLSIGGDPFMD--HPFDVKWELTANNVIMVSAIGNDGPL--YG 344

DB 209 IVSGIEWATNTGMDVINMSLAGSGASTAMKQAVDNAY---ARGVVVVAAGNSGSGNTN 265

QY 345 TLNPNADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVK 404

DB 266 TIGYPAKYDSVAVGAVDSNRSASFSSVGA---EL-----EVMAPGVSIQSTLPGNKY 313

QY 405 GCRALSGTSVASPVVAGAVTLLYSTVKRELVPASMKQALIASARRLPGVNMFEQHG 464

DB 314 NYATPLNGTSMASPHVAGAAALILS---KHPNWTNTQVRSLENTTTLG--SSFYKKG 368

QY 465 KLDL 468

DB 369 LINV 372

RESULT 20

US-10-090-624-31

; Sequence 31, Application US/10090624

; Patent No. US2002013235A1

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/10/090,624

; PRIOR FILING DATE: 2002-03-06

; PRIOR FILING DATE: 1999-12-06

; PRIOR FILING DATE: 1999-12-06

; PRIOR FILING DATE: 1997-06-10

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31

; LENGTH: 382

; TYPE: PRT

; ORGANISM: Bacillus subtilis

US-10-090-624-31

Query Match

Best Local Similarity 31.7%; Pred. No. 6.9e-19;

Matches 98; Conservative 50; Mismatches 118; Indels 43; Gaps 13;

QY 180 HSSRLLRAIPROVAOTLQADVLWQGYTGANTRVAVFTGLSEKHPHFK-----NVKE 233

DB 102 HVAHAYAQSPVPGVQ--IKAPALHSQGYTGSNVKVAVIDSGIDSSHPDLKVAGGASWVPS 160

QY 234 RTNWTNERTLDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRFTNNQVSYTSWFL 290

DB 161 ETN-----PFODNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWII 215

QY 291 DAFNVAILKKIDVLNLSIGGDPFMD--HPFDVKWELTANNVIMVSAIGNDGP--LYGTLN 346

DB 216 NGIEWATANNMDVINNSIGSPSSAALKAADVKA---VASGVVVVAAGNEGTSGSSTV 272

QY 347 NNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRG--GVK 404

DB 273 GYPGKPSVIAVGNVSSNQASFSVGP---EL-----DVMAPGVSIQSTLPGNK 320

QY 405 GCRALSGTSVASPVVAGAVTLLYSTVKRELVPASMKQALIASARRLPGVNMFEQHG 464

DB 321 YG--AYNGTSMASPHVAGAAALILS---KHPNWTNTQVRSLENTTTLG--DSFYKKG 373

QY 465 KLDLRAYQ 473

DB 374 LINVQAAQ 382

RESULT 21

US-09-813-408-19

; Sequence 19, Application US/09813408

; Publication NO. US20030049619A1

; GENERAL INFORMATION:

; APPLICANT: Delagrave, Simon

; APPLICANT: Maïres, Barry

; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides And Combinatorial

; FILE REFERENCE: HER0041

; CURRENT APPLICATION NUMBER: US/09/813,408

; CURRENT FILING DATE: 2001-03-21

; NUMBER OF SEQ ID NOS: 85

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 19

; LENGTH: 380

; TYPE: PRT

; ORGANISM: Bacillus amyloliquefaciens

US-09-813-408-19

Query Match

Best Local Similarity 31.3%; Pred. No. 7.6e-19;

Matches 96; Conservative 52; Mismatches 119; Indels 40; Gaps 12;

QY 180 HSSRLLRAIPROVAOTLQADVLWQGYTGANTRVAVFTGLSEKHPHFK-----NVKE 233

DB 101 HVAHAYAQSPVPGVQ--IKAPALHSQGYTGSNVKVAVIDSGIDSSHPDLKVAGGASWVPS 159

QY 234 RTNWTNERTLDGLGHGTFVAGVIASMR-----COGFAPDAELHIFRFTNNQVSYTSWFL 290

DB 160 ETN-----PFODNNSHGTHVAGTVAALNNSIGVLGVAPSASLYAVKVLGADGSGQYSWII 214

QY 291 DAFNVAILKKIDVLNLSIGGDPFMDHPFPVVKWELTANNVIMVSAIGNDGP--LYGTLN 348

DB 215 NGIEWATANNMDVINNSIGSPSSAALKAADVKA---VASGVVVVAAGNEGTSGSSTV 272

QY 349 PADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRG--GVKGG 406

DB 273 PGKPSVIAVGNVSSNQASFSVGP---EL-----DVMAPGVSIQSTLPGNKY 320

QY 407 CEALSGTSVASPVVAGAVTLLYSTVKRELVPASMKQALIASARRLPGVNMFEQHGKL 466

DB 321 --AYNGTSMASPHVAGAAALILS---KHPNWTNTQVRSLENTTTLG--DSFYKKG 373

```
Qy 467 DLLRAYQ 473
Db 374 NVQAAAQ 380

RESULT 22
US-09-837-235-17
; Sequence 17, Application US/09837235
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Marshall, Christopher
; APPLICANT: Hoffman, Alexander
; APPLICANT: Errico, Joseph
; APPLICANT: Marshall, Paul
; TITLE OF INVENTION: STABILIZED PROTEINS
; FILE REFERENCE: 9725-005-999
; CURRENT APPLICATION NUMBER: US/09/837,235
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28595
; PRIOR FILING DATE: 2000-10-16
; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 17
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-837-235-17

Query Match 5.7%; Score 318.5; DB 10; Length 275;
Best Local Similarity 30.7%; Pred. No. 5e-19;
Matches 91; Conservative 55; Mismatches 119; Indels 31; Gaps 11;

Qy 187 RATPROVAOTLQADVLQMGYTGANVRVAVFDTGLSEKHHPHFNKERTNWTNERT--LD 244
Db 2 QSVFYGISQ-IPALHSQGTGTSNVKVAVIDSGIDSHDPL-NVRGASFVPSETNPYQ 59
;
245 DGLGHGTFVAGVTASMR--COGFAPDAELHIFRVFTNNQVSTSWFLDAFNAILKKI 301
Db 60 DGSNGTHVAGTTAALNNSIGLVGPSASLYAVKVLDSGTSGQYSWIINGEWAISNM 119
;
302 DVNLSTGGPDFMD--HPFVDKVMELTANNVIMVSAIGNDGP--LYGTLLNPADQMDVIG 357
Db 120 DVNLSLGGPTGTALKTVVDKA---VSSGIVVAAAAGNEGSGSTSTVGYPAKYPSTIA 176
;
358 VGGIDFEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVKGGRALSGTSVAS 417
Db 177 VGAVNSNQASFSAGS---EL-----DVNAPGVSIQSTLPGGTYGAYNGTOMAT 224
;
418 PVVAGAVTLVSTVQKRELNVNPMKQALIASARLPGVNMFEQGHGKLDLLRAYQ 473
Db 225 PHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--NSFYVGKGLINVQAAAQ 275
;

RESULT 23
US-09-813-408-17
; Sequence 17, Application US/09813408
; Publication No. US20030049619A1
; GENERAL INFORMATION:
; APPLICANT: Delagrave, Simon
; APPLICANT: Maris, Barry
; TITLE OF INVENTION: Methods For The Synthesis Of Polynucleotides
; TITLE OF INVENTION: Of Polynucleotides
; FILE REFERENCE: HER0041
; CURRENT APPLICATION NUMBER: US/09/813,408
; CURRENT FILING DATE: 2001-03-21
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 17
; LENGTH: 274
; TYPE: PRT
```

```
; ORGANISM: Bacillus subtilis
US-09-813-408-17

Query Match 5.6%; Score 317; DB 9; Length 274;
Best Local Similarity 30.8%; Pred. No. 6.6e-19;
Matches 91; Conservative 54; Mismatches 120; Indels 30; Gaps 11;

Qy 187 RATPROVAOTLQADVLQMGYTGANVRVAVFDTGLSEKHHPHFNKERTNWTNERT--LD 244
Db 2 QSVFYGISQ-IPALHSQGTGTSNVKVAVIDSGIDSHDPL-NVRGASFVPSETNPYQ 59
;
245 DGLGHGTFVAGVTASMR--COGFAPDAELHIFRVFTNNQVSTSWFLDAFNAILKKI 302
Db 60 DGSNGTHVAGTTAALNNSIGLVGPSASLYAVKVLDSGTSGQYSWIINGEWAISNM 119
;
303 VLNLISGGPDFMD--HPFVDKVMELTANNVIMVSAIGNDGP--LYGTLLNPADQMDVIG 358
Db 120 VNNLSLGGPTGTALKTVVDKA---VSSGIVVAAAAGNEGSGSTSTVGYPAKYPSTIAV 176
;
359 GGIDFEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVKGGRALSGTSVASP 418
Db 177 GAVNSNQASFSAGS---EL-----DVNAPGVSIQSTLPGGTYGAYNGTOMATP 224
;
419 VVAGAVTLVSTVQKRELNVNPMKQALIASARLPGVNMFEQGHGKLDLLRAYQ 473
Db 225 HVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--NSFYVGKGLINVQAAAQ 274
;

RESULT 24
US-10-090-624-12
; Sequence 12, Application US/10090624
; Patent No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match 5.6%; Score 316; DB 12; Length 659;
Best Local Similarity 26.8%; Pred. No. 3.3e-18;
Matches 126; Conservative 65; Mismatches 192; Indels 88; Gaps 21;

Qy 199 ADVLM-QMGYTGANVRVAVFDTGLSEKHHPHFNKERTNWTNERT--TNRITLDDGLGHGTFV 253
Db 147 ADTVNLSLGGPTGTALKTVVDKA---VSSGIVVAAAAGNEGSGSTSTVGYPAKYPSTIAV 204
;
254 AGVTASMR--COGFAPDAELHIFRVFTNNQVSTSWFLDAFNAILKKI-----IDVLN 305
Db 205 AGIVAGTGSVNSQYIGVAPGAKLVGVKVLADSGSGSVSTIAGVDVWVQNKDKGIRVIN 264
;
306 LSTG-----GPDFMDHPFVDKVMELTANNVIMVSAIGNDGPPLYTLNPNADQMDVIGV 359
Db 265 LSLGSSQSSDGTSLQA-VNNAWDA---GIVVCAAGNSGPNITYTVGSPAASAKVITVG 320
;
360 GIDFEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGV---RGSVKGKGC-----RAL 410
Db 321 AVDSNDNIASFSSRG-----PTADRLKPEVAVPGVDIIAPRASGTSMTPTINDYTTKA 374
;
```

116 ANNDVIMNLSGPGSSAALAAAYDKA---VAGVVVVAAGNEGTSSTVGPYKYP 172
354 DVICVGGIDPEDNIARFSSRGMTTWELPGGYGEMKPDIVTYGAGVRS--GVKGGCRALS 411
173 SVIAGVAVDSSNQASFSVGP---EL-----DVMAPOVSIQSTLPGNKYG--AYN 218
412 CTSVASPVVAGAVTLLVSTVOKRELVNPAASKOALIASARRLPVGNMFEQCHGKLDLIRA 471
219 GTSMAHPVAGAAALILS---KHPNWTINTQVRSLENTTKLG---DSFYKGLINVOAA 273
472 YQ 473
274 AQ 275
Search completed: May 29, 2003, 13:41:25
Job time : 56 secs

411 SGTSVASPVVAGAVTLLVSTVOKRELVNPAASKOALIASARRLPVGNMFE--QHGKLDL 468
375 SGTSMATPHVSGVALI---LQAHPSWTDPKVKTALETADIAPKEIADIAGAGRVN 431
469 LRAVIGLLNSKPOASL--SPSYIDLTECPYMPYCSQPIYGGMPVWVNTILNGMVTG 526
432 YKAKI---YDDYAKLTFTGSAADKGSATHTDVSG-----AUFVTATLYWDTG--- 476
527 RIVKDPWQDYL-PONGDNIEVAES-----YSSVLPWPSGYLAISISVTKKAASW 575
477 -----SSDIDLILDPNGNEVDYSYTAIFYGKGVYNFTAGTW-----TKVVSY 522
576 EGIAGCHVMITVASPAETESKNGAEQSTVKLPKIKVLIPTPPSKRYLWD 626
523 KGAANYQDV-VSDGSLSGSGGNPNPNPNETPTTDTQTFTGTVNDYWD 572

RESULT 25
US-08-322-678-7
; Sequence 7, Application US/08322678
; Publication No. US20030077807A1
; GENERAL INFORMATION:
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genencor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Horn, Margaret A.
; REGISTRATION NUMBER: 33,401
; REFERENCE/DOCKET NUMBER: GC235-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 742-7536
; TELEFAX: (415) 742-7217
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-678-7

Query Match 5.6%; Score 314.5; DB 1; Length 275;
Best Local Similarity 32.1%; Pred. No. 1.le-18;
Matches 97; Conservative 49; Mismatches 113; Indels 43; Gaps 13;
QY 187 RAYPROVAGTQADVLQMGYTGANVRVAVDTGLSEKHPFK-----NVKERTWNE 240
Db 2 QSPVYQVQ- IKAPALHSQGYTGSNVKVAVIDSGIDSHDPLKAVAGASWYPSFN---- 56
QY 241 RTLDDGLGHGHPVAGVTASMR-----COGFAPDAELHIFVFTNNQVSYTSMFLDAFNAL 297
Db 57 -PFQDNNSHGTHVAGTVAAALNSIGVLGVAFPSASLYAVKVLGADSGSGYSWINGIEWAI 115
QY 298 LKKIDVLNLSIGGDFEMD--HPFVDKVMELTANNVIMVSAIGNDGP--LYGTLLNFDQM 353

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model.

Run on: May 29, 2003, 13:35:17 ; Search time 29 Seconds
(Without alignments)
1067.342 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIMLLLVLLVLCGRKH.....PRYKRPQLMQVHPKTPSV 1052

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Maximum Match 100%

Listing first 45 summaries

Database:

1: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*
3: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*
4: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*
5: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*
6: /cgnt2_6/prodata/1/1aa/5A.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5617	100.0	1052	US-09-255-502-7	Sequence 7, Appli
2	5617	100.0	1052	US-09-360-237-3	Sequence 3, Appli
3	5487	97.7	1052	US-09-360-237-1	Sequence 1, Appli
4	338.5	6.0	352	US-07-923-260A-6	Sequence 6, Appli
5	337.5	6.0	352	US-07-923-260A-1	Sequence 1, Appli
6	337.5	6.0	381	US-07-772-087-5	Sequence 5, Appli
7	337.5	6.0	381	US-08-322-965-2	Sequence 2, Appli
8	330.5	5.9	275	US-07-864-298-1	Sequence 1, Appli
9	329.5	5.9	379	US-08-845-295A-1	Sequence 1, Appli
10	329.5	5.9	379	US-09-140-933-1	Sequence 1, Appli
11	329.5	5.9	379	US-09-146-661-1	Sequence 1, Appli
12	329.5	5.9	379	US-09-150-515-1	Sequence 1, Appli
13	329.5	5.9	350	US-07-923-260A-4	Sequence 4, Appli
14	327	5.8	378	US-07-772-087-4	Sequence 4, Appli
15	323.5	5.8	275	US-08-322-677A-8	Sequence 8, Appli
16	323.5	5.8	275	US-08-322-676-8	Sequence 8, Appli
17	323.5	5.8	275	US-08-898-218-8	Sequence 8, Appli
18	323.5	5.8	275	US-08-848-793-8	Sequence 8, Appli
19	323.5	5.8	275	US-09-255-502-3	Sequence 3, Appli
20	323.5	5.8	275	US-09-178-155-4	Sequence 4, Appli
21	323.5	5.8	275	US-09-445-270-3	Sequence 3, Appli
22	323.5	5.8	275	US-09-445-270-3	Sequence 3, Appli
23	323.5	5.8	352	US-07-923-260A-3	Sequence 3, Appli
24	322	5.7	382	US-09-178-155-2	Sequence 2, Appli
25	321.5	5.7	382	US-09-178-155-2	Sequence 2, Appli
26	320	5.7	734	US-09-000-016-4	Sequence 4, Appli
27	320	5.7	734	US-09-514-340-4	Sequence 4, Appli

28	320	5.7	823	4	US-09-000-016-2	Sequence 2, Appli
29	320	5.7	823	4	US-09-514-340-2	Sequence 2, Appli
30	319.5	5.7	382	2	US-08-504-265B-75	Sequence 75, Appli
31	319.5	5.7	382	4	US-09-445-472-31	Sequence 31, Appli
32	318.5	5.7	352	1	US-07-923-260A-2	Sequence 2, Appli
33	318.5	5.7	382	1	US-08-460-343B-2	Sequence 2, Appli
34	318.5	5.7	382	1	US-08-398-028B-2	Sequence 2, Appli
35	318.5	5.7	382	2	US-08-504-265B-2	Sequence 2, Appli
36	318.5	5.7	382	4	US-09-255-502-2	Sequence 2, Appli
37	316.5	5.6	379	1	US-08-685-774-2	Sequence 2, Appli
38	316.5	5.6	659	4	US-08-894-818B-5	Sequence 5, Appli
39	316	5.6	381	1	US-08-173-508-12	Sequence 12, Appli
40	316	5.6	381	2	US-08-265-310-12	Sequence 12, Appli
41	316	5.6	381	3	US-08-951-742-12	Sequence 12, Appli
42	316	5.6	659	4	US-08-894-818B-1	Sequence 1, Appli
43	316	5.6	659	4	US-09-445-472-12	Sequence 12, Appli
44	315.5	5.6	659	4	US-09-000-016-7	Sequence 7, Appli
45	315.5	5.6	520	4	US-09-514-340-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1	US-09-255-502-7	Application US/09255502
Sequence 7, Appli	US-09-255-502-7	Patent No. 6218165
GENERAL INFORMATION:		
APPLICANT:	Estell, David	
TITLE OF INVENTION:	Harding, Fiona	
TITLE OF INVENTION:	Mutant Proteins Having Lower Allergenic Responses in Humans and Methods for Constructing, Identifying and Producing Such Proteins	
FILE REFERENCE:	GC 527-D2	
CURRENT FILING DATE:	US/09/255,502	
PRIOR FILING DATE:	1999-02-23	
NUMBER OF SEQ ID NOS:	7	
SOFTWARE:	PatentIn Ver. 2.0	
SEQ ID NO 7		
LENGTH:	1052	
TYPE:	PRT	
ORGANISM:	Homo sapiens	
US-09-255-502-7		
Query Match	100.0%; Score 5617; DB 4; Length 1052;	
Best Local Similarity	100.0%; Pred. No. 0;	
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
Db	1	1
1	MKLVNIMLLLVLLVLCGRKHLDRLKSKSEKAPCCSHLTKVERSSVVEEYIAF	60
1	MKLVNIMLLLVLLVLCGRKHLDRLKSKSEKAPCCSHLTKVERSSVVEEYIAF	60
61	NGFTAKARNSFSSALKSEEDVNRITPRNPSDDPSPFEVQIEKOKAGLLTEDEH	120
61	NGFTAKARNSFSSALKSEEDVNRITPRNPSDDPSPFEVQIEKOKAGLLTEDEH	120
121	PNIKRYTPQKRVRSKLTAEADPTPCNETRWQKQSSRPRLRASLSGFWHATGRH	180
121	PNIKRYTPQKRVRSKLTAEADPTPCNETRWQKQSSRPRLRASLSGFWHATGRH	180
181	SSRRLRAIRPROVAQTQADVLMQMGYTGANVRVAVDTGLSEKHPFKVKEKRTNTNE	240
181	SSRRLRAIRPROVAQTQADVLMQMGYTGANVRVAVDTGLSEKHPFKVKEKRTNTNE	240
241	RTIDDDGCHGTAVAGVYASMRCEGFPADDELHFRFTNNQVSYTSWFLDAFVATLTK	300
241	RTIDDDGCHGTAVAGVYASMRCEGFPADDELHFRFTNNQVSYTSWFLDAFVATLTK	300
301	IDVLNLSIGGPDPMDFVYKRVMLANNTYNSAIGNDPLXGTLNPNADQMDVIGVG	360
301	IDVLNLSIGGPDPMDFVYKRVMLANNTYNSAIGNDPLXGTLNPNADQMDVIGVG	360

Mon Jun 2 11:46:23 2003

us-09-830-837-6.ra1

Page 2

QY 361 IDFDNINARSSKGMTTWELPGYGRMKPDIVYTGAVGSGVKGGRALSGTSVASPVY 420
DB 361 IDFDNINARSSKGMTTWELPGYGRMKPDIVYTGAVGSGVKGGRALSGTSVASPVY 420
QY 421 AGAVTLVSTVOKRELVPNSKMOALIASARLPGVNMFGHGLDLRLAYOLINSYK 480
DB 421 AGAVTLVSTVOKRELVPNSKMOALIASARLPGVNMFGHGLDLRLAYOLINSYK 480
QY 481 QASISPSYIDLTECPYMPYCOPITYYGMPTVVNTILNGMVGRIYDKPDMQPTLPQ 540
DB 481 QASISPSYIDLTECPYMPYCOPITYYGMPTVVNTILNGMVGRIYDKPDMQPTLPQ 540
QY 541 NGDNIEVAFSYSVLMPWSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
DB 541 NGDNIEVAFSYSVLMPWSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
QY 601 QSTVVKLPITKVKIIPPPRSKRYLMDQYHNLRYPGYFPDNLRMKNDPLDMNGDHIHTN 660
DB 601 QSTVVKLPITKVKIIPPPRSKRYLMDQYHNLRYPGYFPDNLRMKNDPLDMNGDHIHTN 660
QY 661 FPDWYOHLSMGYFVEVLGAPFCFASOYGTLLMVDSEEEYPEREIAKLARDVDNGLSL 720
DB 661 FPDWYOHLSMGYFVEVLGAPFCFASOYGTLLMVDSEEEYPEREIAKLARDVDNGLSL 720
QY 721 VIFSDWYNTSVARKYFYDENTROMMMPDTGGANI PALNELLSVNMGFSOGLYEGEFTL 780
DB 721 VIFSDWYNTSVARKYFYDENTROMMMPDTGGANI PALNELLSVNMGFSOGLYEGEFTL 780
QY 781 ANHDMYASGGSIAKPEDGVVITGOFKDOGLEVLKOFVAVENVYILGLYOIPABEGGR 840
DB 781 ANHDMYASGGSIAKPEDGVVITGOFKDOGLEVLKOFVAVENVYILGLYOIPABEGGR 840
QY 841 IYVGDSCNCLDSHROKCFWLLDALLOYTSYGVTPPSLSHSNRRPRPSGAGSVTERM 900
DB 841 IYVGDSCNCLDSHROKCFWLLDALLOYTSYGVTPPSLSHSNRRPRPSGAGSVTERM 900
QY 901 EGNHLHRYSKYLEAHLGDPKRPRLPACPRLSNAKPOPLNETAPSNLMKQKLLSIDDKY 960
DB 901 EGNHLHRYSKYLEAHLGDPKRPRLPACPRLSNAKPOPLNETAPSNLMKQKLLSIDDKY 960
QY 961 VLPNFRSNRPQVRPLSPGESGAMDIPGIMPGRYNOEVGQTIIVFAFLGAMVYLAFFVYQ 1020
DB 961 VLPNFRSNRPQVRPLSPGESGAMDIPGIMPGRYNOEVGQTIIVFAFLGAMVYLAFFVYQ 1020
QY 1021 INKASRRPRRRKPRVKRPOLMQOVHPKTPSV 1052
DB 1021 INKASRRPRRRKPRVKRPOLMQOVHPKTPSV 1052

RESULT 2
US-09-360-237-3
Sequence 3, Application US/09360237
Feature No. 632962

GENERAL INFORMATION
APPLICANT: BROWN, MICHAEL S.
APPLICANT: CHENG, DONG
APPLICANT: ESPENSHADE, PETER J.
APPLICANT: GOLDSTEIN, JOSEPH L.
APPLICANT: RAMSON, ROBERT B.
APPLICANT: SAKAI, JURO
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
FILE REFERENCE: UTXD:567
CURRENT APPLICATION NUMBER: US/09/360,237
EARLIER FILING DATE: 1999-07-23
EARLIER APPLICATION NUMBER: 60/096,571
NUMBER OF SEQ ID NOS: 60
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 1052
TYPE: PRT

ORGANISM: Human
US-09-360-237-3
Query Match 100.0%; Score 5617; DB 4; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKVNTMLLVLLVLLGCKRHLDREKSPKAPCCGSHLTKYESSTVYEXYVAF 60
DB 1 MKVNTMLLVLLVLLGCKRHLDREKSPKAPCCGSHLTKYESSTVYEXYVAF 60
QY 61 NGFTAKARNSFISALKSEVDNRRITPRNNSSDYPSDFEYQIKKQKAGLLLEH 120
DB 61 NGFTAKARNSFISALKSEVDNRRITPRNNSSDYPSDFEYQIKKQKAGLLLEH 120
QY 121 PNTRKVTPOKRVFSLKYASDPTVPCNETRMSQWOSRPLRASLSLSGFWHTGHN 180
DB 121 PNTRKVTPOKRVFSLKYASDPTVPCNETRMSQWOSRPLRASLSLSGFWHTGHN 180
QY 181 SSRRLLRAIPROVADTLADVLMQMGYGANRVAVEDTGLSEKHPHKYKERTWINE 240
DB 181 SSRRLLRAIPROVADTLADVLMQMGYGANRVAVEDTGLSEKHPHKYKERTWINE 240
QY 241 RTLDDELGHGTFVAGVIAASMRGCGFAPDAELHFRVFTNNVSYTSWFLDAFNVALIK 300
DB 241 RTLDDELGHGTFVAGVIAASMRGCGFAPDAELHFRVFTNNVSYTSWFLDAFNVALIK 300
QY 301 IDVLNLSIGGPFDMHFVDKWEELFANNVIMVSAIGNDPLYLGLLNNPADQMDVIGVG 360
DB 301 IDVLNLSIGGPFDMHFVDKWEELFANNVIMVSAIGNDPLYLGLLNNPADQMDVIGVG 360
QY 361 IDPEDNINARSSKGMTTWELPGYGRMKPDIVYTGAVGSGVKGGRALSGTSVASPVY 420
DB 361 IDPEDNINARSSKGMTTWELPGYGRMKPDIVYTGAVGSGVKGGRALSGTSVASPVY 420
QY 421 AGAVTLVSTVOKRELVPNSKMOALIASARLPGVNMFGHGLDLRLAYOLINSYK 480
DB 421 AGAVTLVSTVOKRELVPNSKMOALIASARLPGVNMFGHGLDLRLAYOLINSYK 480
QY 481 QASISPSYIDLTECPYMPYCOPITYYGMPTVVNTILNGMVGRIYDKPDMQPTLPQ 540
DB 481 QASISPSYIDLTECPYMPYCOPITYYGMPTVVNTILNGMVGRIYDKPDMQPTLPQ 540
QY 541 NGDNIEVAFSYSVLMPWSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
DB 541 NGDNIEVAFSYSVLMPWSGYLAISVTKKAASMEGIAOGHMTIVASPAETESKNGAE 600
QY 601 QSTVVKLPITKVKIIPPPRSKRYLMDQYHNLRYPGYFPDNLRMKNDPLDMNGDHIHTN 660
DB 601 QSTVVKLPITKVKIIPPPRSKRYLMDQYHNLRYPGYFPDNLRMKNDPLDMNGDHIHTN 660
QY 661 FPDWYOHLSMGYFVEVLGAPFCFASOYGTLLMVDSEEEYPEREIAKLARDVDNGLSL 720
DB 661 FPDWYOHLSMGYFVEVLGAPFCFASOYGTLLMVDSEEEYPEREIAKLARDVDNGLSL 720
QY 721 VIFSDWYNTSVARKYFYDENTROMMMPDTGGANI PALNELLSVNMGFSOGLYEGEFTL 780
DB 721 VIFSDWYNTSVARKYFYDENTROMMMPDTGGANI PALNELLSVNMGFSOGLYEGEFTL 780
QY 781 ANHDMYASGGSIAKPEDGVVITGOFKDOGLEVLKOFVAVENVYILGLYOIPABEGGR 840
DB 781 ANHDMYASGGSIAKPEDGVVITGOFKDOGLEVLKOFVAVENVYILGLYOIPABEGGR 840
QY 841 IYVGDSCNCLDSHROKCFWLLDALLOYTSYGVTPPSLSHSNRRPRPSGAGSVTERM 900
DB 841 IYVGDSCNCLDSHROKCFWLLDALLOYTSYGVTPPSLSHSNRRPRPSGAGSVTERM 900
QY 901 EGNHLHRYSKYLEAHLGDPKRPRLPACPRLSNAKPOPLNETAPSNLMKQKLLSIDDKY 960
DB 901 EGNHLHRYSKYLEAHLGDPKRPRLPACPRLSNAKPOPLNETAPSNLMKQKLLSIDDKY 960
QY 961 VLPNFRSNRPQVRPLSPGESGAMDIPGIMPGRYNOEVGQTIIVFAFLGAMVYLAFFVYQ 1020
DB 961 VLPNFRSNRPQVRPLSPGESGAMDIPGIMPGRYNOEVGQTIIVFAFLGAMVYLAFFVYQ 1020

Db 961 VLPNFRSRNPQVRPLSPGESGAMDPGIMGRYNOEVGOTIPVAFIAGAVYLAFFVQ 1020
QY 1021 INKAKSRPKRRKPRVKRPOLMOQVHPKPTSV 1052
Db 1021 INKAKSRPKRRKPRVKRPOLMOQVHPKPTSV 1052

RESULT 3

US-09-360-237-1
Sequence 1, Application US/09360237
Patent No. 6322962
GENERAL INFORMATION:
APPLICANT: BROWN, MICHAEL S.
APPLICANT: CHENG, DONG
APPLICANT: ESPENSHADE, PETER J.
APPLICANT: GOLDSTEIN, JOSEPH L.
APPLICANT: RAMSON, ROBERT B.
APPLICANT: SAKAI, JURO
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
TIME OF INVENTION: MODULATORS THEREOF
FILE REFERENCE: 07XD:567
CURRENT APPLICATION NUMBER: US/09/360, 237
CURRENT FILING DATE: 1999-07-23
EARLIER APPLICATION NUMBER: 60/096, 571
EARLIER FILING DATE: 1998-08-14
NUMBER OF SEQ ID NOS: 60
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 1
LENGTH: 1052
TYPE: PRT
ORGANISM: Cricetus griseus
US-09-360-237-1

Query Match 97.7%; Score 5487; DB 4; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLVNIWLLLVLLVLLCGKHHGLGRLKSEFKAPCPGCSHLLTKVESSVVEYEYIAF 60
Db 1 MKLVNIWLLLVLLVLLCGKHHGLGRLKSEFKAPCPGCSHLLTKVESSVVEYEYIAF 60
QY 61 NGFTAKARNSFTSSALKSEVDNWRITIPNNPSSDYPSPFVTOIKEROKAGILLTEDEH 120
Db 61 NGFTAKARNSFTSSALKSEVDNWRITIPNNPSSDYPSPFVTOIKEROKAGILLTEDEH 120
QY 121 PNIRKRYTPQKRVRSLSKTAASDPYPCNETRMSQKMOSSRLKRAASLSLGSGFHNAIGH 180
Db 121 PNIRKRYTPQKRVRSLSKTAASDPYPCNETRMSQKMOSSRLKRAASLSLGSGFHNAIGH 180
QY 181 SSRRLKAIPTROVAQTQADVYLMOMGTGANVRAVFDTGISEKHPRKVKERTNMTNE 240
Db 181 SSRRLKAIPTROVAQTQADVYLMOMGTGANVRAVFDTGISEKHPRKVKERTNMTNE 240
QY 241 RTLDGIGHGTFAVAGVIAKRECOGAPDAELHIFRYFTNNQVSYTSMFLDAFNATILKK 300
Db 241 RTLDGIGHGTFAVAGVIAKRECOGAPDAELHIFRYFTNNQVSYTSMFLDAFNATILKK 300
QY 301 IDVLNLSIGGPDHDFVOKVWELLTANNYIMSAIGNDGLYGLTLNPPADOMVYIGVG 360
Db 301 IDVLNLSIGGPDHDFVOKVWELLTANNYIMSAIGNDGLYGLTLNPPADOMVYIGVG 360
QY 361 IDEEDNITARRSSRGMWTWELPGYGRMKPDIVTGYAGVRSYGSGGCRALSGTVAAPV 420
Db 361 IDEEDNITARRSSRGMWTWELPGYGRMKPDIVTGYAGVRSYGSGGCRALSGTVAAPV 420
QY 421 AGAVTLIVSVQKRELVNPAKMAQALIASARLPGVNMFEQGHGKLDLRLAYQILNSYKP 480
Db 421 AGAVTLIVSVQKRELVNPAKMAQALIASARLPGVNMFEQGHGKLDLRLAYQILNSYKP 480
QY 481 QASISPSYIDLTETCPYMPYCQPIYYGGMPTVNVITLNGMGTGRIVKPRDQPIYQ 540
Db 481 QASISPSYIDLTETCPYMPYCQPIYYGGMPTVNVITLNGMGTGRIVKPRDQPIYQ 540

QY 541 NGDNIEAFSSYSVLMWPGSLAISISVTKKAASMEGIAQHVMIIVASPAETESKNGAE 600
Db 541 NGDNIEAFSSYSVLMWPGSLAISISVTKKAASMEGIAQHVMIIVASPAETESKNGAE 600
QY 601 QSTVKLPKIVKILPTPPRSKRVLMQYHNLRYPGYPPRDNLKMKNDPLDMNGDIHTN 660
Db 601 HTSTVKLPKIVKILPTPPRSKRVLMQYHNLRYPGYPPRDNLKMKNDPLDMNGDIHTN 660
QY 661 FRDMYQHLRSMGVVEVLGAPFTCEPDASQYGTLLMNDSEEEYFPEETAKLRDNDGLSL 720
Db 661 FRDMYQHLRSMGVVEVLGAPFTCEPDASQYGTLLMNDSEEEYFPEETAKLRDNDGLSL 720
QY 721 VIFSDWNTSVMRKVEYDETRMMPDTCGANIPALNELLSYNNMGFSGLYEGEFTL 780
Db 721 VIFSDWNTSVMRKVEYDETRMMPDTCGANIPALNELLSYNNMGFSGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGVITQTERDQLEVLKQETAVENVPIGLYQIPAEAGGR 840
Db 781 ANHDMYASGCSIAKFPEDGVITQTERDQLEVLKQETAVENVPIGLYQIPAEAGGR 840
QY 841 IYLVGDSNCLDSDHROKDFWLDLALQYTSYGYTPPSLSHGNORPPSGAGSYTERM 900
Db 841 IYLVGDSNCLDSDHROKDFWLDLALQYTSYGYTPPSLSHGNORPPSGAGSYTERM 900
QY 901 EGNHLHRSKYLEAHLGDPKPRPLPACPRLSMAKPOPENETAPSNLTKHOKLSIDLDKY 960
Db 901 EGNHLHRSKYLEAHLGDPKPRPLPACPRLSMAKPOPENETAPSNLTKHOKLSIDLDKY 960
QY 961 VLPNFRSRNPQVRPLSPGESGAMDPGIMGRYNOEVGOTIPVAFIAGAVYLAFFVQ 1020
Db 961 VLPNFRSRNPQVRPLSPGESGAMDPGIMGRYNOEVGOTIPVAFIAGAVYLAFFVQ 1020
QY 1021 INKAKSRPKRRKPRVKRPOLMOQVHPKPTSV 1052
Db 1021 INKAKSRPKRRKPRVKRPOLMOQVHPKPTSV 1052

RESULT 4

US-07-923-260A-6
Sequence 6, Application US/07923260A
Patent No. 5719021
GENERAL INFORMATION:
APPLICANT: Inouye, Masaori
TITLE OF INVENTION: PROTEIN ACTIVATION
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gerard J. Weiser
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/923, 260A
FILING DATE: 31-JUL-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Weiser, Gerard J.
REGISTRATION NUMBER: 19,763
REFERENCE/DOCKET NUMBER: 377,5638P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-875-8383
TELEFAX: 215-875-8394
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 352 amino acids
TYPE: amino acid
TOPOLOGY: linear

RESULT 6

US-07-772-087-5
Sequence 5, Application US/07772087
Patent No. 5275945
GENERAL INFORMATION:
APPLICANT: HSIAO, Hung-Yu
APPLICANT: FODGE, Douglas W.
APPLICANT: LALONDE, James J.
TITLE OF INVENTION: ALKALINE PROTEASES STABLE IN HEAVY-DUTY
TITLE OF INVENTION: DETERGENT LIQUIDS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
City: Alexandria
STATE: Virginia
COUNTRY: USA
ZIP: 2213-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/772,087
FILING DATE: 19911008
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16754/115 CHCO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 381 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus
STRAIN: B. subtilis
US-07-772-087-5

Query Match 6.0%; Score 337.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 4.6e-19;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVEVEEYVAVNGFTAKARNSFTSSALKSSSEVDNWKRIIPRNPSSDIPSPDEVIQIK 107
DB 33 SST-EKKTYVGF-----KOTMSAMSSAKKKDYI----- 59
QY 108 EKOKAGILTLTE-DHPNIKRYTPQRKYFRSLKYAESDPTVPCNETRMSQKQSSRPLRRAS 166
DB 60 -SEKGGVQKQFYVNAANAATLDEKAVKELK---KDPSSVA----- 95
QY 167 LSLGSGFWATGHRSSRRLLRAIPROYAQTLDVLMQMGYTGANVAVFDTGLSEKHP 226
DB 96 -----YEEEDHIAHEYVQSVYGISQ- IKAPALHSQGYTGSNVAVAIDSGIDSHP 146
QY 227 HFKNVKERTWMTNERT--LDDGLGHGTFVAGVIASME---COGFADALHIFRYFTNN 281
DB 147 DL-NVRGASFEVSETPYQDSSGHGTHVAGTIALNNSIGVLGVSFASLAVKYLVDST 205
QY 282 QVSTWTFDLAFNVAIILKIDVNLSTIGPDFMD--HPFDKWEIETANNVIMVSAIGND 339
DB 206 GSGQYWIINGIEMIALSNMNDVIMNSLGGPTGSTALTAKTVVDKA---VSSGIIVAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVIGVGIGIDEDNIARFSSRQMTWELPGVGRKKPDIYTVGAG 397
DB 263 GSSGSGTSTVGYPAKYPTSTIAVGAVNSSNQASFSAGS---EL-----DVMARGVS 310

QY 398 VRSGVVGCCRALSGTISVAPVAGAVTLLVSTYQKREIYNPASMOKALLISARRLPGVN 457
DB 311 IQSTLPQGTGYGAVTSMATPHVAGAAALILS---KHPTWTNAOVRDLSESTAYYLQ--N 365
QY 458 MEEQHGKLDLLRAYQ 473
DB 366 SFYGGKGLINVOAAQ 381

RESULT 7

US-08-322-965-2
Sequence 2, Application US/08322965
Patent No. 5733473
GENERAL INFORMATION:
APPLICANT: Johnston, James P.
APPLICANT: Lenoir, Pierre M.A.
APPLICANT: Thoen, Christian A.J.K.
TITLE OF INVENTION: Liquid detergent Composition containing
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 5299 Spring Grove Avenue
City: Cincinnati
STATE: Ohio
COUNTRY: USA
ZIP: 45217
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,965
FILING DATE: 13-OCT-1994
CLASSIFICATION: 252
ATTORNEY/AGENT INFORMATION:
NAME: Allen, George W.
REGISTRATION NUMBER: 26,143
REFERENCE/DOCKET NUMBER: CM-352MC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 513/627-5946
TELEFAX: 513/627-8118
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 381 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-322-965-2

Query Match 6.0%; Score 337.5; DB 1; Length 381;
Best Local Similarity 26.8%; Pred. No. 4.6e-19;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVEVEEYVAVNGFTAKARNSFTSSALKSSSEVDNWKRIIPRNPSSDIPSPDEVIQIK 107
DB 33 SST-EKKTYVGF-----KOTMSAMSSAKKKDYI----- 59
QY 108 EKOKAGILTLTE-DHPNIKRYTPQRKYFRSLKYAESDPTVPCNETRMSQKQSSRPLRRAS 166
DB 60 -SEKGGVQKQFYVNAANAATLDEKAVKELK---KDPSSVA----- 95
QY 167 LSLGSGFWATGHRSSRRLLRAIPROYAQTLDVLMQMGYTGANVAVFDTGLSEKHP 226
DB 96 -----YEEEDHIAHEYVQSVYGISQ- IKAPALHSQGYTGSNVAVAIDSGIDSHP 146
QY 227 HFKNVKERTWMTNERT--LDDGLGHGTFVAGVIASME---COGFADALHIFRYFTNN 281
DB 147 DL-NVRGASFEVSETPYQDSSGHGTHVAGTIALNNSIGVLGVSFASLAVKYLVDST 205
QY 282 QVSTWTFDLAFNVAIILKIDVNLSTIGPDFMD--HPFDKWEIETANNVIMVSAIGND 339

Db 206 GSGQSVITINGIENAMISNNNDVIMNLSGPGTSTALKTVDKA---VSSGIYVAAAANE 262
Oy 340 GP--LYGLINPADOMOVIGVIGDIPEDNIARFSSRGKTTWELPGYGRKPDYTGAG 397
Db 263 GSSGSIYGPAPKPTSTIAGVAVNSNORASFSSAGS--EL-----DYAPGV 310
Oy 398 VRSQVKGCCRALSTGVASPVYAGVTLVSTVOKRELVPASMKOALIASARRLPVY 457
Db 311 IOSTLPGGYKINGTSMATPHVAGAAALITS--KHPTWNAOVRDLSTATYLG--N 365
Oy 458 MFEQHGKIDLRAYQ 473
Db 366 SFYGGKGLINVOAAQ 381

RESULT 8
US-07-864-298-1
Sequence 1, Application US/07864298
Patent No. 5316935
GENERAL INFORMATION:
APPLICANT: Arnold, Frances H.
TITLE OF INVENTION: No. 5316935el Subtilisin Variants
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harris F. Brotman, POMS, SMITH, LANDE & ROSE
STREET: 2121 Avenue of the Stars Suite 1400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOOS
SOFTWARE: Patent release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/864, 298
FILING DATE: 1992/04/06
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Brotman, Harris F.
REGISTRATION NUMBER: 35,461
REFERENCE/DOCKET NUMBER: 91-124
TELECOMMUNICATION INFORMATION:
TELEPHONE: (310) 788-5000
TELEFAX: (310) 277-1297
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus subtilis
STRAIN: 1168
US-07-864-298-1

Query Match 5.9%; Score 330.5; DB 1; Length 275;
Best Local Similarity 31.8%; Pred. No. 1e-18;
Matches 94; Conservative 54; Mismatches 117; Indels 31; Gaps 11;
Oy 187 RAIPROAVOTLADVLMQMGYTGANVRAVDTGLSEKHPFNKERTNMTNERT--LD 244
Db 2 GSVPGISQ--IKAPALHSOGYTSNKKVAVIDSGIDSHDL--NVGGSFVPSSTNRYQ 59
Oy 245 DSIGHGTAVAGVIAISME--CGFAPDAELHIFRFTNNQVSYTSWFLDAENYALIKKI 301
Db 60 DSSHGTHVAGTIALNNSIGVGLVAPASALYAVKVLDTSGQSVIINGIENAMISNM 119

Oy 302 DYVNLISIGGDFEMD--HPPVDKWEILANNVIVSAIGNDGP--LYGTINPADOMDYG 357
Db 120 DYVNLISIGGFTGTALKTYVDKA---VSSGIYVAAAANEBSGSIYSTGYAPKPTSTIA 176
Oy 358 VGGIDFEDNIARFSSRGKTTWELPGYGRKPDYTGAGYVRSQVKGCCRALSTGVAS 417
Db 177 VGANVSNORASFSSAGS--EL-----DYAPGVSIQSTLPGGYKINGTSMAT 224
Oy 418 PVYAGVATLVSTVOKRELVPASMKOALIASARRLPVNMFEQHGKIDLRAYQ 473
Db 225 PHVAGAAALITS--KHPTWNAOVRDLSTATYLG--NSFYGGKGLINVOAAQ 275

RESULT 9
US-08-845-295A-1
Sequence 1, Application US/08845295A
Patent No. 5817490
GENERAL INFORMATION:
APPLICANT: Hubbs, John C.
TITLE OF INVENTION: Enzymatic Process for the Manufacture of
TITLE OF INVENTION: Ascorbic Acid, 2-Keto-L-Gulonic Acid, and Esters of
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eastman Chemical Company
STREET: P. O. Box 511
CITY: Kingsport
STATE: Tennessee
COUNTRY: USA
ZIP: 37662-5075
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/845, 295A
FILING DATE: 25-April-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017, 879
FILING DATE: 17-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: Cheryl J. Tubach
REGISTRATION NUMBER: 38,346
REFERENCE/DOCKET NUMBER: 70432
TELECOMMUNICATION INFORMATION:
TELEPHONE: 423-229-6189
TELEFAX: 423-229-1239
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-845-295A-1

Query Match 5.9%; Score 329.5; DB 2; Length 379;
Best Local Similarity 25.3%; Pred. No. 2e-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;
Oy 1 MKLVINWLLLVLLCGKKHIGDLREKKSFERAPCGCSHLTKVERFSTV-----V 52
Db 2 MRKSPWLGTLAFM-----LVETMAFSDSASAAQAPKRV 36
Oy 53 EYEYIAF--NGYFTAKANSPFISALKSSEVD--NMRLIPRNNSDYPSPDEYVIOIREKO 110
Db 37 EKDYIGFSGVGYKVDIKES--GKVDKOPRIINAkakALD----- 80
Oy 111 KAGLTLEHPNINIKVTYPOKRVFSLKYAESDPYPCNERTFMSQKMSRPLRASISLG 170
Db 81 KEALKEVKNDPPV-----AYVEED----- 99

```

QY 171 SGFMATGRSSRLRLAPROVAOTLADLYLWOMQYTAANVAEFDLSKHFPEKN 230
Db 100 -----HAHALAAGTVPYCI-PLIKADYKQAGCFKANKYVALDITGICASHEDLVN 149
QY 231 VKERTNMTNERTLDGLGHGTFFVAGVIASMRCCQ--GAPDAELHIFRVTNNQVSYS 287
Db 150 VGASFEVAGEAYNTDGNCGHGTFFVAGTVAALDMTTGLVGAAPSVLTAAYVNLSSGSGTYS 209
QY 288 WFLDAFNAILKIDIVLNLISIGPPBMD--HPFVDVWELTANVNLVMSAIGNDPL- Y 343
Db 210 GIVSIEBATTNGMIVNINSLGSPGSGTLMKQAVNDAY---ARGVYVVAALAGSGSGNT 266
QY 344 GTLNNPADQMDVYIGVGIDFEDNLIAFSSRGMTTWELPGYGMRKPDIVTYGAGVRGSGV 403
Db 267 NTIGYPAYDYSLVIAAGVADSNSNRASFSSVGA---EL-----EVNAPGAGVYSTYP 314
QY 404 KGGCRALSGTSVASFVAVGAVYLLVSTVQKRREIVNPASMKOALIASARLLPGVNMFEQGH 463
Db 315 TSTVATLNGTSMASPHVAGAAALITLS--KHBNILMSQVRRNRLSSATYTLG--SSFYLYK 369
QY 464 GKLDLLRAYQ 473
Db 370 GLINVEAAAO 379

```

[illegible]

Mon Jun 2 11:46:23 2003

us-09-830-837-6.rai

Page 8

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: Protein
US-09-146-661-1

Query Match 5.9%; Score 329.5; DB 4; Length 379;
Best Local Similarity 25.3%; Pred. No. 26-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

1 MKLVNIMLLVLLVLLCGKKHLDRLKSKFEKAPCGCSHLTLKVEFSSTV-----V 52
2 MKKSPFWGLMLTAFM-----LVFTMAFSDSASAAQPAKAV 36
3 EYEYIAF-NGEYTAARNSTFSSALKSEVD-NMRTIPRNPSSDYPDEYIOIKKQ 110
4 EKDYIVGKSGVTAASKDILKES--GGKVDKQRIINAKAKLD-----80
5 37
6 81 KEAKKEKNDPV-----AYEED-----99
7 171 SGFWATGRHSSRLRLAIPROVAOTLQADVLMQGTGANVAVAFDTGLSEKHPEKN 230
8 100 -----HVAHALAQVTPYGI-PLIKADKVOAGFGANVAVAVLDIGIASHPDLV 149
9 231 VKERTNMTNERTLDGLGHTGTFVAGVIAASMRCC--GFAPDELHIFRYFTNNQVSYTS 287
10 150 VGGASVAGAEVNTDNGHGTHTVAGTVALDNTTGVLDGAPSVLAAYVLLSSSGSYTS 209
11 288 WFLDAFNATILKKIDVNLISIGDPFMD--HPVDKWEITANNVIMVSAIGNDPTL--Y 343
12 210 GIVSGIEMATTNGMDVYINMSIGPSSGTAMKQAVDNAY--ARGVYVAAAGNSSSGNT 266
13 344 GTLNNPADOMDVIGVIGIDFEDNIARFSSRGMTWELPGYGRMKPDITYGAGVRSQV 403
14 267 NTIGYPAKYDSYIANGVANDSNRASFSSVGA--EL-----EVAAPAGYSTYP 314
15 404 KGGCRALSGTSVAPVAVGATLLVSTVQKRELVPASMKCALIASRRLPVMNFEQGH 463
16 315 TSYTATLNTGTSMAHPVAGAAALILS--KHNLNLSAQSQRNRLSSTATYLG--SSFYGK 369
17 464 GKLDLIRAYO 473
18 370 GLINVEAAQ 379

RESULT 12
US-09-150-515-1
Sequence 1, Application US/09150515
Patent No. 6271006
GENERAL INFORMATION:
APPLICANT: Hubbs, John C.
TITLE OF INVENTION: Enzymatic Process for the Manufacture of
TITLE OF INVENTION: Ascorbic Acid, 2-Keto-L-Gulonic Acid, and Esters of
TITLE OF INVENTION: 2-Keto-L-Gulonic Acid
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Eastman Chemical Company
STREET: P.O. Box 511
CITY: Kingsport
STATE: Tennessee
COUNTRY: USA
ZIP: 37662-5075
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/150, 515

FILING DATE: 09-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/017,879; 08/845,295
FILING DATE: 17-May-96; 25-April-97
ATTORNEY/AGENT INFORMATION:
NAME: Cheryl J. Tubach
REGISTRATION NUMBER: 38,346
REFERENCE/DOCKET NUMBER: 70432
TELECOMMUNICATION INFORMATION:
TELEPHONE: 423-229-6189
TELEFAX: 423-229-1239
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 379 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
US-09-150-515-1

Query Match 5.9%; Score 329.5; DB 4; Length 379;
Best Local Similarity 25.3%; Pred. No. 26-18;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;

1 MKLVNIMLLVLLVLLCGKKHLDRLKSKFEKAPCGCSHLTLKVEFSSTV-----V 52
2 MKKSPFWGLMLTAFM-----LVFTMAFSDSASAAQPAKAV 36
3 EYEYIAF-NGEYTAARNSTFSSALKSEVD-NMRTIPRNPSSDYPDEYIOIKKQ 110
4 EKDYIVGKSGVTAASKDILKES--GGKVDKQRIINAKAKLD-----80
5 37
6 81 KEAKKEKNDPV-----AYEED-----99
7 171 SGFWATGRHSSRLRLAIPROVAOTLQADVLMQGTGANVAVAFDTGLSEKHPEKN 230
8 100 -----HVAHALAQVTPYGI-PLIKADKVOAGFGANVAVAVLDIGIASHPDLV 149
9 231 VKERTNMTNERTLDGLGHTGTFVAGVIAASMRCC--GFAPDELHIFRYFTNNQVSYTS 287
10 150 VGGASVAGAEVNTDNGHGTHTVAGTVALDNTTGVLDGAPSVLAAYVLLSSSGSYTS 209
11 288 WFLDAFNATILKKIDVNLISIGDPFMD--HPVDKWEITANNVIMVSAIGNDPTL--Y 343
12 210 GIVSGIEMATTNGMDVYINMSIGPSSGTAMKQAVDNAY--ARGVYVAAAGNSSSGNT 266
13 344 GTLNNPADOMDVIGVIGIDFEDNIARFSSRGMTWELPGYGRMKPDITYGAGVRSQV 403
14 267 NTIGYPAKYDSYIANGVANDSNRASFSSVGA--EL-----EVAAPAGYSTYP 314
15 404 KGGCRALSGTSVAPVAVGATLLVSTVQKRELVPASMKCALIASRRLPVMNFEQGH 463
16 315 TSYTATLNTGTSMAHPVAGAAALILS--KHNLNLSAQSQRNRLSSTATYLG--SSFYGK 369
17 464 GKLDLIRAYO 473
18 370 GLINVEAAQ 379

RESULT 13
US-07-923-260A-4
Sequence 4, Application US/07923260A
Patent No. 5719021
GENERAL INFORMATION:
APPLICANT: Inouye, Masayori
TITLE OF INVENTION: PROTEIN ACTIVATION
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Gerard J. Weiser
STREET: 230 South Fifteenth Street, Suite 500
CITY: Philadelphia

OY 429 STVOKRELVNPMASKQALLIASARRLPVNMFEQHGKIDLLRAVQ 473
DB 339 S---KHPNLSASOVNRNLSSTATYLG--SSFYGGKGLINVEAAAO 378

RESULT 15

US-08-322-677A-8
; Sequence 8, Application US/08322677A
; Patent No. 5677272
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Burns, Michael E.
; APPLICANT: Digulio, David N.
; APPLICANT: Getty, Edward E.
; APPLICANT: Hartshorn, Richard T.
; APPLICANT: Willey, Alan D., III
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Bleaching Compositions Comprising Protease Enzymes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,677A
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William
; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5041R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-677A-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
OY 187 RAIPROVAQTLQADVLMQMGYTGANRVAVFDITGLSEKHPHFKNVKERTNMTNERT--LD 244
DB 2 OSVPYGISQ-IRAPALHSQGYTGSNVKAVAVIDGSDSHPD-LNVGGSASFPSPTNPYQ 59
OY 245 DGLGHTFVAGVIAEMRE--COGFAPDAELHIFVFTNNQVSTFMPFADANVAILKKI 301
DB 60 DSSHGTHVAGTIALNNSIGVLGVSASLYAVAVLDSTGSGQYSMIINGIEMAINNM 119
OY 302 DVYNTSIGSPDFMD--HPFVDKVMELTANNVIMVSAIGNDP--LVGTLANPADQMDVIG 357
DB 120 DVYNTSIGSPDTALTKTVYDKA---VSSGIYVAAAGNBSGSGTSTYGYAPAKPSTIA 176
OY 358 VGGIDPEFNINARSSRGMTTWELPGGYGNMKPDYTYTGAGYRSGYKGGCRALSGTYSAS 417
DB 177 VGVAVNSNORASFSASG--EL-----DVNAPEVSTISQTLPGGTYGAYNKGTSMT 224

OY 418 PVVAGAVTLVSTVOKRELVNPMASKQALLIASARRLPVNMFEQHGKIDLLRAVQ 473
DB 225 PVVAGAAALITIS---KHPNLSASOVNRNLSSTATYLG--NSFYGGKGLINVOAAAO 275

RESULT 16

US-08-322-676-8
; Sequence 8, Application US/08322676
; Patent No. 5679630
; GENERAL INFORMATION:
; APPLICANT: Baeck, Andre (NMN)
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Graycar, Thomas P.
; APPLICANT: Bott, Richard R.
; APPLICANT: Wilson, Lori J.
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Protease-Containing Cleaning Compositions
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,676
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William
; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5040R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 275 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-322-676-8

Query Match 5.8%; Score 323.5; DB 1; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;
OY 187 RAIPROVAQTLQADVLMQMGYTGANRVAVFDITGLSEKHPHFKNVKERTNMTNERT--LD 244
DB 2 OSVPYGISQ-IRAPALHSQGYTGSNVKAVAVIDGSDSHPD-LNVGGSASFPSPTNPYQ 59
OY 245 DGLGHTFVAGVIAEMRE--COGFAPDAELHIFVFTNNQVSTFMPFADANVAILKKI 301
DB 60 DSSHGTHVAGTIALNNSIGVLGVSASLYAVAVLDSTGSGQYSMIINGIEMAINNM 119
OY 302 DVYNTSIGSPDFMD--HPFVDKVMELTANNVIMVSAIGNDP--LVGTLANPADQMDVIG 357
DB 120 DVYNTSIGSPDTALTKTVYDKA---VSSGIYVAAAGNBSGSGTSTYGYAPAKPSTIA 176
OY 358 VGGIDPEFNINARSSRGMTTWELPGGYGNMKPDYTYTGAGYRSGYKGGCRALSGTYSAS 417
DB 177 VGVAVNSNORASFSASG--EL-----DVNAPEVSTISQTLPGGTYGAYNKGTSMT 224

Mon Jun 2 11:46:23 2003

us-09-830-837-6.ra1

Page 12

OY 358 VGGIDFEDNINARSSRGMTWELPGYGRMKPDIYVYAGVSGVKGCGALSGTSVAS 417
Db 177 VGVANSSNQRAFSFSSAGS--EL-----DWMAPVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARRLPGVMMFPOGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWTNAQVRDLRESTATYILG--NSFYGGKGLINVOAAQ 275

RESULT 19
US-09-255-502-3
Sequence 3, Application US/09255502
Patent No. 6218165

GENERAL INFORMATION:
APPLICANT: Estell, David
APPLICANT: Harding, Fiona
TITLE OF INVENTION: Mutant Proteins Having Lower Allergenic Responses in
TITLE OF INVENTION: Humans and Methods for Constructing, Identifying and
TITLE OF INVENTION: Producing Such Proteins
FILE REFERENCE: GC 527-D2
CURRENT FILING DATE: 1998-02-23
PRIOR FILING DATE: 1998-04-15
PRIORITY NUMBER: 09/060,872
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 3
LENGTH: 275
ORGANISM: Bacillus subtilis
US-09-255-502-3

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVACTLOADVLQMGVYGANVAVFDTGLSEKHPKKNVKEKNTNERT--LD 244
Db 2 QSVYVIGSQ--IKAPLHSGGYTSNKKVAVIDSGIDSHPD--NVRGASVPSSETNPYQ 59
OY 245 DGLGHGTFFVAGVIASMR--COGFAPDAELHIFVFTNNVSYTSWFLDAFNVAIILKRI 301
Db 60 DGSCHGTAVAGTIALNNSIGVGPSASLTVAVLDSTGSGYVILINGIEMAISNM 119
OY 302 DVNLSTIGGPDND--HPEVDKWEITANNVIVSAIGNDP--LYGTLPNPDQMDVYG 357
Db 120 DVNMSLIGGPTGTALKTVDKA---VSSGIYVAAAAGNBSGSGSTGYGPAKYPSTIA 176
OY 358 VGGIDFEDNINARSSRGMTWELPGYGRMKPDIYVYAGVSGVKGCGALSGTSVAS 417
Db 177 VGVANSSNQRAFSFSSAGS--EL-----DWMAPVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARRLPGVMMFPOGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWTNAQVRDLRESTATYILG--NSFYGGKGLINVOAAQ 275

RESULT 20
US-09-178-155-4
Sequence 4, Application US/09178155
Patent No. 6312936
GENERAL INFORMATION:
APPLICANT: Poulos, Aycookaran J.
APPLICANT: Schellenderger, Volker
APPLICANT: Kellis, Jr., James T.
APPLICANT: Paech, Christian
APPLICANT: Naehery, Joanne
APPLICANT: Naki, Donald P.
TITLE OF INVENTION: Multiply-Substituted Protease Variants
FILE REFERENCE: GC502-2
CURRENT FILING DATE: US/09/178,155
CURRENT FILING DATE: 1998-10-23

EARLIER APPLICATION NUMBER: 08/956,323
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,564
EARLIER FILING DATE: 1997-10-23
EARLIER APPLICATION NUMBER: 08/956,324
EARLIER FILING DATE: 1997-10-23
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 275
TYPE: PRT
ORGANISM: B. subtilis
US-09-178-155-4

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

OY 187 RAIPROVACTLOADVLQMGVYGANVAVFDTGLSEKHPKKNVKEKNTNERT--LD 244
Db 2 QSVYVIGSQ--IKAPLHSGGYTSNKKVAVIDSGIDSHPD--NVRGASVPSSETNPYQ 59
OY 245 DGLGHGTFFVAGVIASMR--COGFAPDAELHIFVFTNNVSYTSWFLDAFNVAIILKRI 301
Db 60 DGSCHGTAVAGTIALNNSIGVGPSASLTVAVLDSTGSGYVILINGIEMAISNM 119
OY 302 DVNLSTIGGPDND--HPEVDKWEITANNVIVSAIGNDP--LYGTLPNPDQMDVYG 357
Db 120 DVNMSLIGGPTGTALKTVDKA---VSSGIYVAAAAGNBSGSGSTGYGPAKYPSTIA 176
OY 358 VGGIDFEDNINARSSRGMTWELPGYGRMKPDIYVYAGVSGVKGCGALSGTSVAS 417
Db 177 VGVANSSNQRAFSFSSAGS--EL-----DWMAPVSIQSTLPGGTGAYNGTSMAT 224
OY 418 PVVAGAVTLVSTVQKRELVPASMKOALIASARRLPGVMMFPOGHGKLDLIRAYQ 473
Db 225 PVVAGAAALILS---KHPTWTNAQVRDLRESTATYILG--NSFYGGKGLINVOAAQ 275

RESULT 21
US-09-445-270-3
Sequence 3, Application US/09445270
Patent No. 6369011
GENERAL INFORMATION:
APPLICANT: Rai, Saroj
APPLICANT: Correa, Paul
APPLICANT: Zhu, Yong
APPLICANT: Graycar, Thomas
APPLICANT: Bolt, Richard
TITLE OF INVENTION: Protease Enzymes for Tough Cleaning and
TITLE OF INVENTION: Compositions Incorporating Same
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: The Procter & Gamble Company
STREET: 11810 East Miami River Road
CITY: Ross
STATE: OH
COUNTRY: USA
ZIP: 45061
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA
APPLICATION NUMBER: US/09/445,270
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Bolam, Brian M.
REGISTRATION NUMBER: 37,513
REFERENCE/DOCKET NUMBER: 1234
TELECOMMUNICATION INFORMATION:

TELEPHONE: 513/627-2457
TELEFAX: 513/627-0318
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 275 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-445-270-3

Query Match 5.8%; Score 323.5; DB 4; Length 275;
Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

QY 187 RAIPROYAOTLQADVLMQGYTGANRVAVFDGLSEKHFFKNVKEKRTWTERT--LD 244
DB 2 QSVYGISO--IKAPALHSOGYTGSNKVAVIDSGIDSSHDL--NVRGASFEVSETPYQ 59
QY 245 DGLGHGTFFVAGVIAASRE--COGFAPDAELHIFRVETNNQVSYTSMFLDAFVYALIKKI 301
DB 60 DSSSHGTHVAGTTAALNNSIGVLGVSPASLIVAKYLDSTGSCQYSMLINGIEMALISNM 119
QY 302 DVNLNLSIGGPDMD--HPFVYKWEELTANNVIVSAIGNDGP--LYGTLLNPPADQMDVIG 357
DB 120 DVINMSLGGPTGTALKTVDKA--VSSGIYVAAAAGNCGSSGSTVGYPAKYPSTIA 176
QY 358 VGGIDEDNINARSSKGMTWELPGYGRMKPDIVYVYAGVRSQVKGCGRAISGTSVAS 417
DB 177 VGVVNSSNORASSSAGS--EL-----DVMAPGVSIQSTLPGITGAYGANGTSMAT 224
QY 418 PVVAGAVTLVSTVOKRELVPNPSMKOALIASARLPVGMFPOGHGKLDLLRAYQ 473
DB 225 PVVAGAAALILS--KHPTWNAQVRDLRESTATYIG--NSFYGGKGLINVOAAQ 275

RESULT 22

5472855-4

Patent No. 5472855

APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.

TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS

NUMBER OF SEQUENCES: 31

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/287,964

FILING DATE: 22-SEP-1994

Prior Application DATA:

APPLICATION NUMBER: 90,902

FILING DATE: 12-JUL-1993

APPLICATION NUMBER: 823,039

FILING DATE: 14-JAN-1992

APPLICATION NUMBER: 35,652

FILING DATE: 06-APR-1987

APPLICATION NUMBER: 334,081

FILING DATE: 04-APR-1989

APPLICATION NUMBER: 127,134

FILING DATE: 01-DEC-1987

APPLICATION NUMBER: 846,627

FILING DATE: 01-APR-1986

APPLICATION NUMBER: 858,594

FILING DATE: 30-APR-1986

APPLICATION NUMBER: 614,612

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,615

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,617

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,491

FILING DATE: 29-MAY-1984

SEQ ID NO: 4:

LENGTH: 275

Query Match

5.8%; Score 323.5; DB 6; Length 275;

Best Local Similarity 31.1%; Pred. No. 3.7e-18;
Matches 92; Conservative 55; Mismatches 118; Indels 31; Gaps 11;

QY 187 RAIPROYAOTLQADVLMQGYTGANRVAVFDGLSEKHFFKNVKEKRTWTERT--LD 244
DB 2 QSVYGISO--IKAPALHSOGYTGSNKVAVIDSGIDSSHDL--NVRGASFEVSETPYQ 59
QY 245 DGLGHGTFFVAGVIAASRE--COGFAPDAELHIFRVETNNQVSYTSMFLDAFVYALIKKI 301
DB 60 DSSSHGTHVAGTTAALNNSIGVLGVSPASLIVAKYLDSTGSCQYSMLINGIEMALISNM 119
QY 302 DVNLNLSIGGPDMD--HPFVYKWEELTANNVIVSAIGNDGP--LYGTLLNPPADQMDVIG 357
DB 120 DVINMSLGGPTGTALKTVDKA--VSSGIYVAAAAGNCGSSGSTVGYPAKYPSTIA 176
QY 358 VGGIDEDNINARSSKGMTWELPGYGRMKPDIVYVYAGVRSQVKGCGRAISGTSVAS 417
DB 177 VGVVNSSNORASSSAGS--EL-----DVMAPGVSIQSTLPGITGAYGANGTSMAT 224
QY 418 PVVAGAVTLVSTVOKRELVPNPSMKOALIASARLPVGMFPOGHGKLDLLRAYQ 473
DB 225 PVVAGAAALILS--KHPTWNAQVRDLRESTATYIG--NSFYGGKGLINVOAAQ 275

RESULT 23

US-07-923-260A-3

Sequence 3, Application US/07923260A

Patent No. 5719021

GENERAL INFORMATION:

APPLICANT: Inouye, Masayori

TITLE OF INVENTION: PROTEIN ACTIVATION

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSER: Gerard J. Welser

STREET: 230 South Fifteenth Street, Suite 500

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/923,260A

FILING DATE: 31-JUL-1992

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Welser, Gerard J.

REGISTRATION NUMBER: 19,763

REFERENCE/DOCKET NUMBER: 377,563BP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 215-875-8363

TELEFAX: 215-875-8394

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 352 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Bacillus amylosacchariticus

FEATURE:

NAME/KEY: Region

LOCATION: 1..77

OTHER INFORMATION: /note- "The region from 1 to 77 is

Query Match

5.8%; Score 323.5; DB 1; Length 352;

Mon Jun 2 11:46:23 2003

us-09-830-837-6.rai

Page 15

```

QY 347 NNPDADQMDYIGVGIDFEDNINARFSSRGHTTWELPGYGRMPDIYTYAGAVNGS--GYK 404
Db 273 GYPGRYPYVIAAGVADSSNQRASFSFVGP--EL-----DVNAPGVSIOSTLP;GKK 320
QY 405 GGCRAISGTSVAPYVAGAVTLTSTVQKRELVPNASMKQALIASARLLPGVMNEPQHG 464
Db 321 YG--AYNGISMSPHVAGAAALILS--KHPMWTNTQVRSSILENTTTLKLG--DSFYFGKG 373
QY 465 KIDLRLAYQ 473
Db 374 LINVQAAAO 382

```

Search completed: May 29, 2003, 13:40:25
Job time : 32 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:27:27 ; Search time 75 Seconds
(without alignments)
1869.062 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MLLVNIWLLVLLGKKH.....PRVKRQLMQVHPPTPSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5617	100.0	1052	21	Human subtilisin-k
2	5617	100.0	1052	21	Amino acid sequenc
3	5617	100.0	1052	21	Human subtilisin.
4	5617	100.0	1052	21	Human subtilisin p
5	5617	100.0	1052	22	Human subtilisin
6	5617	100.0	1052	23	Human site-1 prote
7	5487	97.7	1052	21	Amino acid sequenc
8	5487	97.7	1052	23	Hamster site-1 pro
9	5476	97.5	1052	21	Rat subtilisin-kex
10	5427	96.6	1052	21	Mouse subtilisin-k

11	3549	63.2	666	23	ABB90255	Human polypeptide
12	2366.5	42.1	992	22	AB220015	Drosophila site 1
13	2231	39.7	952	22	AB220015	Drosophila melanog
14	528	9.4	126	22	AA001619	Human polypeptide
15	340.5	6.1	381	7	AA00571	Sequence of subtil
16	340.5	6.1	381	11	AA007970	Sequence of Bacill
17	340.5	6.1	381	11	AA007970	Subtilisin gene pro
18	338.5	6.0	352	19	AA046600	Suppressor mutatio
19	337.5	6.0	352	19	AA046595	Amino acid sequenc
20	337.5	6.0	381	8	AA071060	Wild-type subtilis
21	337.5	6.0	381	11	AA03737	Subtilisin gene pr
22	337.5	6.0	381	13	AA024131	Bacterial serine p
23	337.5	6.0	381	14	AA034463	Bacillus subtilis
24	337.5	6.0	381	16	AA074224	B. subtilis subtili
25	337.5	6.0	381	20	AA039229	Bacillus subtilis
26	336.5	6.0	381	8	AA070053	Bacillus subtilis
27	336.5	6.0	1079	22	AA081180	Transglutaminase r
28	335.5	6.0	379	23	AA019062	Bacillus lichenifo
29	333.5	5.9	275	15	AA051928	Bacillus subtilis
30	332.5	5.9	381	23	AA019063	Bacillus subtilis
31	330.5	5.9	275	15	AA051920	Bacillus subtilis
32	330.5	5.9	275	15	AA051924	Bacillus subtilis
33	330.5	5.9	275	15	AA051929	Bacillus subtilis
34	330.5	5.9	275	15	AA052643	Bacillus subtilis
35	329.5	5.9	275	15	AA051922	Bacillus subtilis
36	329.5	5.9	275	15	AA051923	Bacillus subtilis
37	329.5	5.9	275	15	AA051926	Bacillus subtilis
38	329.5	5.9	275	15	AA051927	Bacillus subtilis
39	329.5	5.9	379	19	AA031600	Subtilisin useful
40	329	5.9	350	19	AA046598	Amino acid sequenc
41	327.5	5.8	275	12	AA010207	Mutant subtilisin
42	327.5	5.8	275	15	AA052644	B. subtilis subtili
43	326.5	5.8	275	12	AA010206	Mutant subtilisin
44	326.5	5.8	275	12	AA010208	Mutant subtilisin
45	326.5	5.8	275	12	AA010205	Mutant subtilisin

ALIGNMENTS

RESULT 1
AAB06334 AAB06334 standard; Protein; 1052 AA.

XX AC AAB06334;

DT 03-OCT-2000 (first entry)

XX Human subtilisin-kexin isoenzyme 1.

XX Human; subtilisin-kexin isoenzyme 1; SKI-1; antilipemic; cytostatic;
XX vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
XX Ras-dependent cancer; restenosis; amyloid protein formation;
XX pro-brain-derived neurotrophic factor; proBDNF;
XX sterol-regulatory element-binding protein; SREBP.
XX Homo sapiens.

PN WO200026348-A2.

XX 11-MAY-2000.

PF 04-NOV-1999; 99WO-CA01058.

PR 04-NOV-1998; 98CA-2249648.

PA (RECL-) INST RECH CLINIQUES MONTREAL.

PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;

XX WPI; 2000-365601/31.

DR N-PSDB; AAA57197.

XX

PT Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 PT producing a polypeptide useful for treating hypercholesterolemia, liver
 PT steatosis and amyloidosis, comprises a specific amino acid sequence -
 XX
 PS
 XX
 Claim 1; Page 56-59; 119pp; English.
 CC The present sequence is human subtilisin-kexin isoenzyme 1 (SKI-1),
 CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
 CC from the human neuronal cell line IMR-32 by RT-PCR using active site
 CC degenerate primers. SKI-1 cleaves at a specific threonine residue within
 CC the N-terminal segment of human pro-brain-derived neurotrophic factor
 CC (proBDNF). It is also capable of cleaving steryl-regulatory
 CC element-binding proteins (SREBPs), which function to control lipid
 CC biosynthesis and uptake in animal cells. Peptides which bind to and are
 CC treated by SKI-1 may be used for monitoring SKI-1 activity, for screening
 CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
 CC activity. Proteic fragments of SKI-1 which bind to the SKI-1
 CC catalytic site may be used as inhibitors of SKI-1 activity. They may
 CC be used to treat diseases involving overexpression of SKI-1 or SKI-1 of
 CC substrate. Such diseases include hypercholesterolemia, high levels of
 CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
 CC Ras-dependent cancer, restenosis and amyloid protein formation.
 XX
 SQ Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNIWLLLVLLCGKKHLDRLKSKPEKAPCGCSHLTLKVFSSVTVVEYIVAF 60
 DB 1 MKLVNIWLLLVLLCGKKHLDRLKSKPEKAPCGCSHLTLKVFSSVTVVEYIVAF 60
 QY 61 NGYFTAKARNSISSALKSSEVDNWRITPRNPPSSDPSDFEVIQIKKOKAGLLTLEDH 120
 DB 61 NGYFTAKARNSISSALKSSEVDNWRITPRNPPSSDPSDFEVIQIKKOKAGLLTLEDH 120
 QY 121 PNKVTQPKVRSLSKYAESDTPVPCNETRWQSKQSRPLRASLSLGGFWHATGRH 180
 DB 121 PNKVTQPKVRSLSKYAESDTPVPCNETRWQSKQSRPLRASLSLGGFWHATGRH 180
 QY 181 SRRLLRAIPQVAQTLQADVLQMGVYTGANRVAVDFGLSEKHPFKKVKERTNWTNE 240
 DB 181 SRRLLRAIPQVAQTLQADVLQMGVYTGANRVAVDFGLSEKHPFKKVKERTNWTNE 240
 QY 241 RLDDGLGHGTFVAGVIASMKRECQGFAPDAELHIFRVFTNNQVSYTSWFLDAENYAILKK 300
 DB 241 RLDDGLGHGTFVAGVIASMKRECQGFAPDAELHIFRVFTNNQVSYTSWFLDAENYAILKK 300
 QY 301 IDVLNLSIGGPDMDHFFVDKVMWELTANNVTMVSATGNDGPLYGLTANNPADQMDVIGVGG 360
 DB 301 IDVLNLSIGGPDMDHFFVDKVMWELTANNVTMVSATGNDGPLYGLTANNPADQMDVIGVGG 360
 QY 361 IDFDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPVV 420
 DB 361 IDFDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRGSGVGGCRALSGTSVASPVV 420
 QY 421 AGAVTLVSTVQKRELVPNSMKQALIASARPLPGVNMFEQGHGKLDLLRAYOILNSYKP 480
 DB 421 AGAVTLVSTVQKRELVPNSMKQALIASARPLPGVNMFEQGHGKLDLLRAYOILNSYKP 480
 QY 481 QASLSPSIDTECPYHWPYCSQPIYTGGMPTVNTILNGMVTGRIVDKPDQOPIYLPQ 540
 DB 481 QASLSPSIDTECPYHWPYCSQPIYTGGMPTVNTILNGMVTGRIVDKPDQOPIYLPQ 540
 QY 541 NGDNIETAFSSVSLWFWGSLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 DB 541 NGDNIETAFSSVSLWFWGSLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 QY 601 QTSTVKLPKIKYIIPTPPKSRKVLWDQYHNLRYPGYPFRDNLKMKNDPLDWDNGDHIHTN 660
 DB 601 QTSTVKLPKIKYIIPTPPKSRKVLWDQYHNLRYPGYPFRDNLKMKNDPLDWDNGDHIHTN 660

QY 661 FDMYQHLSRSMGYFVVLGAPTCFDASQGTGTLMDVDSSEYFPEEIAKLRRDYNGLSL 720
 DB 661 FDMYQHLSRSMGYFVVLGAPTCFDASQGTGTLMDVDSSEYFPEEIAKLRRDYNGLSL 720
 QY 721 VIFSDWYNTSMRKVKFYDENTROWMPDGTGGANIPALNELLISVNMWGFSDGLYEGEFTL 780
 DB 721 VIFSDWYNTSMRKVKFYDENTROWMPDGTGGANIPALNELLISVNMWGFSDGLYEGEFTL 780
 QY 781 ANHDMYASGCSIAKEPFGVYITQTFKQDQGLEVLKQETAVVENPIILGLYQIIPAEGGGR 840
 DB 781 ANHDMYASGCSIAKEPFGVYITQTFKQDQGLEVLKQETAVVENPIILGLYQIIPAEGGGR 840
 QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
 DB 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
 QY 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHQLLSIDLDKV 960
 DB 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHQLLSIDLDKV 960
 QY 961 VLPFRSNRQVRLPSGSGGAWDIPGGIMPGRYNODVGOTIPVFAFLGANVVLAFVWQ 1020
 DB 961 VLPFRSNRQVRLPSGSGGAWDIPGGIMPGRYNODVGOTIPVFAFLGANVVLAFVWQ 1020
 QY 1021 INKAKSRPKRKKPRVVRKPOLMOQVHPKTPSV 1052
 DB 1021 INKAKSRPKRKKPRVVRKPOLMOQVHPKTPSV 1052

RESULT 2
 AAY84227
 ID AAY84227 standard; peptide; 1052 AA.
 AC AAY84227;
 AC AAY84227;
 DT 03-JUL-2000 (first entry)
 DE Amino acid sequence of a human site-1 protease.
 KW Modulator; sterol-regulated site-1 protease; cholesterol;
 KW sterol regulatory element binding protein; SREBP; lipid synthesis;
 KW fatty acid biosynthesis; site-1 protease; protease inhibitor;
 KW serum cholesterol; hypercholesterolemia; lipid metabolism.
 OS Homo sapiens.
 XX WO200009677-A2.
 XX 24-FEB-2000.
 XX 13-AUG-1999; 99WO-US18544.
 XX 14-AUG-1998; 98US-0096571.
 XX 23-JUL-1999; 99US-0360237.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PA (BROW) BROWN M S.
 PA (CHEN) CHENG D.
 PA (ESPE) ESPENSHADE P J.
 PA (GOLD) GOLDSTEIN J L.
 PA (RAWS) RAWSON R B.
 PA (SAKA) SAKAI J.
 XX Brown MS, Cheng D, Espenshade PJ, Goldstein JL, Rawson RB;
 PI Sakai J;
 XX WPI; 2000-224327/19..
 XX Novel assay for identifying modulators of sterol-regulated Site-1
 PT protease useful for the treatment of hypercholesterolemia, involves
 PT identifying an agent capable of down regulating Site-1 protease
 PT activity

PS Claim 62; Fig 4A-B; 172pp; English.

CC The specification describes a method for identifying modulators of a
CC steroid-regulated site-1 protease. Site-1 protease cleaves sterol
CC regulatory element binding proteins (SREBPs) in the endoplasmic
CC reticulum, initiating release from membranes and activating lipid
CC synthesis. The modulators therefore also modulate cholesterol and
CC fatty acid biosynthesis. The method comprises selecting an agent capable
CC of down regulating site-1 protease and formulating a composition
CC comprising the agent. The site-1 protease inhibitors are useful for
CC treating a patient for elevated serum cholesterol. Diseases treated
CC include hypercholesterolemia and other lipid metabolism associated
CC conditions. The present sequence represents a human site-1 protease.

XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYIVAF 60
Db 1 MKLVNWLVLVLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYIVAF 60
QY 61 NGYFTAKARNSFTISSALKSSEVDNWRILIPRNPSSDYPDFEVIQIKKQKAGLLLEHDH 120
Db 61 NGYFTAKARNSFTISSALKSSEVDNWRILIPRNPSSDYPDFEVIQIKKQKAGLLLEHDH 120
QY 121 PNIRKVTPOKVRSLKYAESDPTVPCNETRWSOKWQSSRPLRRASLSLGSFGFWHATGRH 180
Db 121 PNIRKVTPOKVRSLKYAESDPTVPCNETRWSOKWQSSRPLRRASLSLGSFGFWHATGRH 180
QY 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPHFKNVKERTWNTNE 240
Db 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDTGLSEKHPHFKNVKERTWNTNE 240
QY 241 RTLDGGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
Db 241 RTLDGGLGHTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDMPDFVVKVWELTANNVIMVSAIGNDGPLYGLTNLPADQMDVIGVG 360
Db 301 IDVLNLSIGGPDMPDFVVKVWELTANNVIMVSAIGNDGPLYGLTNLPADQMDVIGVG 360
QY 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGRGSGVKGCRALSGTSVASPVV 420
Db 361 IDFDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGRGSGVKGCRALSGTSVASPVV 420
QY 421 AGAVTLVSTVQKRELNVNPSAKQALIASARRLPVNMPEQGHGKLDLLRAYQILNSYRP 480
Db 421 AGAVTLVSTVQKRELNVNPSAKQALIASARRLPVNMPEQGHGKLDLLRAYQILNSYRP 480
QY 481 QASLSPSYIDLTCEPYMPCSQPIYIGMPTVVNTILNGMGTGRIYDKPDQWQYLPQ 540
Db 481 QASLSPSYIDLTCEPYMPCSQPIYIGMPTVVNTILNGMGTGRIYDKPDQWQYLPQ 540
QY 541 NGDNIEVAISYSLVMPWSGYLAISVTKKAASWEGIAQGHVMTFVSPAETESKNGAE 600
Db 541 NGDNIEVAISYSLVMPWSGYLAISVTKKAASWEGIAQGHVMTFVSPAETESKNGAE 600
QY 601 QTSTVKLPKIKIIPPRSKRVLDQYHNLRYPCGYFPRDNLRMKNDPLDNGDHIHN 660
Db 601 QTSTVKLPKIKIIPPRSKRVLDQYHNLRYPCGYFPRDNLRMKNDPLDNGDHIHN 660
QY 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGLLWVDSEEFPEPIAKLRDNDVNGLSL 720
Db 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGLLWVDSEEFPEPIAKLRDNDVNGLSL 720
QY 721 VIFSDWYNTSVMRKVKFYDENTRQWMPDGTGGANIPALNELLISVNMWGSGLYEGEFTL 780
Db 721 VIFSDWYNTSVMRKVKFYDENTRQWMPDGTGGANIPALNELLISVNMWGSGLYEGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGVVITQTFKQDGLVLEKQETAVVENVPILGLYQIPAGGGGR 840

Db 781 ANHDMYASGCSIAKFPEDGVVITQTFKQDGLVLEKQETAVVENVPILGLYQIPAGGGGR 840
QY 841 IVLYGDSNCLDDSHROKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHROKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPERM 900
QY 901 EGNLHRYSKVLEAHILGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
Db 901 EGNLHRYSKVLEAHILGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDKV 960
QY 961 VLPNFSNRQVPRPLSPGSGANDIPGGINPGRYNOVGOTIPVFAFLGAMVLAFTVQ 1020
Db 961 VLPNFSNRQVPRPLSPGSGANDIPGGINPGRYNOVGOTIPVFAFLGAMVLAFTVQ 1020
QY 1021 INKAKSRPRKRKPRVPRPQLMQOVHPKTPSV 1052
Db 1021 INKAKSRPRKRKPRVPRPQLMQOVHPKTPSV 1052

RESULT 3
AAV66982

ID AAY66982 standard; Protein; 1052 AA.

XX AAY66982;

DT 15-MAR-2000 (first entry)

DE Human subtilisin.

XX Subtilisin; allergy; human; bacterium; protease; epitope; detergent;
KW cosmetic; textile; pet food industry; debriement treatment.

XX Homo sapiens.

XX WO953078-A2.

XX 21-OCT-1999.

XX 15-APR-1999; 99WO-US08177.

XX 15-APR-1998; 98US-0060854.

XX (GEMV) GENENCOR INT INC.

XX Estell D;

XX WPI; 2000-061971/05.

XX Modified non-human protease having reduced allergenicity -

XX Disclosure; Fig 6; 38pp; English.

XX This sequence represents the human subtilisin protein. The invention
XX relates to a method of reducing the allergenicity of a non-human
XX protein, especially a bacterial protease such as subtilisin (AAY66980),
XX by identifying an epitope on the non-human protein and replacing it by
XX an analogous region from the human protein. The method is useful for
XX producing proteins, including proteases, that are less likely to cause
XX allergic reactions, thus the invention can be employed for example in
XX detergents, cosmetics, textile treatment, and pet food industries. The
XX human subtilisin can be used pharmaceutically for debriement
XX treatments.

XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 21; Length 1052;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYIVAF 60

Db 1 MKLVNWLVLVLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYIVAF 60

```
QY 61 NGYFTAKARNSFTSSALKSEVDNWRRIIPRNPSSDYPSDFEVIQIKOKAGLLTLEDH 120
DB 61 NGYFTAKARNSFTSSALKSEVDNWRRIIPRNPSSDYPSDFEVIQIKOKAGLLTLEDH 120
QY 121 PNKRVTQPKRFRSLKYAESDPTVPCNETRWQSKWSSRLRRLRASLSLGSFHWATGRH 180
DB 121 PNKRVTQPKRFRSLKYAESDPTVPCNETRWQSKWSSRLRRLRASLSLGSFHWATGRH 180
QY 181 SSRLLRAIPROVAOTLQADVLQWQYTGANVRVAFDPTGLSEKHPKFKVKNKERTNWTNE 240
DB 181 SSRLLRAIPROVAOTLQADVLQWQYTGANVRVAFDPTGLSEKHPKFKVKNKERTNWTNE 240
QY 241 RTLDGLGHGTFVAGVIASMRCEGAPDAELHIFRVFNNOVSYTSWFLDAFNAILKK 300
DB 241 RTLDGLGHGTFVAGVIASMRCEGAPDAELHIFRVFNNOVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDMDHPFVDKRWELTANNVIMVSAIGNDGPLYGTLLNPDQMDVIGVGG 360
DB 301 IDVLNLSIGGPDMDHPFVDKRWELTANNVIMVSAIGNDGPLYGTLLNPDQMDVIGVGG 360
QY 361 IDFDNTARFSSRGMTTWELPGYGRMKPDIVTYGAGVRGSGVKGGRALSGTSVASPVY 420
DB 361 IDFDNTARFSSRGMTTWELPGYGRMKPDIVTYGAGVRGSGVKGGRALSGTSVASPVY 420
QY 421 AGAVTLVSVQKRELVPASMKQALIASARRLPGVNMFEGHKGKLDLLRAYQILNSYKP 480
DB 421 AGAVTLVSVQKRELVPASMKQALIASARRLPGVNMFEGHKGKLDLLRAYQILNSYKP 480
QY 481 QASLSPSYIDLTECPYMPYCSQPIYYGGMTVVNVTILNGMGVTVGRIVDKPDWQPYLPQ 540
DB 481 QASLSPSYIDLTECPYMPYCSQPIYYGGMTVVNVTILNGMGVTVGRIVDKPDWQPYLPQ 540
QY 541 NGDNEVAFSVSSVLPWNSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
DB 541 NGDNEVAFSVSSVLPWNSGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
QY 601 QSTVKLPKIKVIPTPPSKRVLWDQYHNLRYPCYPRNLKRNKNDPLDNGDHIHTN 660
DB 601 QSTVKLPKIKVIPTPPSKRVLWDQYHNLRYPCYPRNLKRNKNDPLDNGDHIHTN 660
QY 661 FRDWOHLRSMGYFEVLGAPFTCDASQYGTLLAVDSEEFEEETAKLRDNDGLSL 720
DB 661 FRDWOHLRSMGYFEVLGAPFTCDASQYGTLLAVDSEEFEEETAKLRDNDGLSL 720
QY 721 VIFSDWNTSVMRKVKFYDENTROWMPDGTGGANTPALNELLSVNNMGFSGGLYGEFTL 780
DB 721 VIFSDWNTSVMRKVKFYDENTROWMPDGTGGANTPALNELLSVNNMGFSGGLYGEFTL 780
QY 781 ANHDMYASGCSIAKFPEDGVVITQTFKDGLEVLKQETAVVENVPTLGLYQIPAEGGGR 840
DB 781 ANHDMYASGCSIAKFPEDGVVITQTFKDGLEVLKQETAVVENVPTLGLYQIPAEGGGR 840
QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
DB 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
QY 901 EGNHLHRYSKVLAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLKV 960
DB 901 EGNHLHRYSKVLAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLKV 960
QY 961 VLPNFRSNRQVRPLSPGESGANDIPGIMPGRYNOYVGTIPVEAFGLNMVYLAFFVVO 1020
DB 961 VLPNFRSNRQVRPLSPGESGANDIPGIMPGRYNOYVGTIPVEAFGLNMVYLAFFVVO 1020
QY 1021 INKAKSPKRRKPRVKRQQLMQVHPKTPSV 1052
DB 1021 INKAKSPKRRKPRVKRQQLMQVHPKTPSV 1052
```

RESULT 4

AA54619

ID AA54619 standard; Protein; 1052 AA.

XX

AC AA54619;

XX 04-FEB-2000 (first entry)

XX Human subtilisin protein sequence.

XX Subtilisin; protease variant; precursor protease; cleaning composition;
detergent; liquid soap application; dish-care formulation; animal feed;
contact lens cleaning solution; peptide hydrolysis; waste treatment;
cosmetic formulation; fusion-cleavage enzyme; protein production.

XX Homo sapiens.

XX WO9953038-A2.

XX 21-OCT-1999.

XX 14-APR-1999; 99WO-US08253.

XX 15-APR-1998; 98US-0060872.

XX (GENV) GENENCOR INT INC.

XX Estell DA, Harding FA;

XX WPI; 2000-013100/01.

XX New mutant proteins having lower allergenic response in humans useful
in cleaning compositions, animal feed and treating textiles -
Example 2; Fig 7; 43pp; English.

XX

XX This sequence represents the human subtilisin. The

XX invention relates to protease variants comprising a substitution at least
1 of the positions in a precursor protease corresponding to K170D, Y171Q
and/or S173D of the Bacillus anyoliquefaciens subtilisin. The protease
variant can be used in a cleaning composition (e.g. in detergents, in bar
or liquid soap application, dish-care formulations, or contact lens
cleaning solutions), in animal feed and for treating textiles
(e.g. treating wool to prevent felting). The protease variant may also be
used for peptide hydrolysis, waste treatment, cosmetic (e.g. skin care),
formulations, or as fusion-cleavage enzymes in protein production. The
proteins (including enzymes) with reduced antigenicity may be used with
peptides which contain epitopes responsible for the individuals exposed.
XX Individual may be identified by measuring the proliferation of T-cells
due to T-cell epitope recognition.

XX Sequence 1052 AA;

XX Query Match 100.0%; Score 5617; DB 21; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNIMLLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYEVAF 60

DB 1 MKLVNIMLLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTVVEYEVAF 60

QY 61 NCYFTAKARNSFTSSALKSEVDNWRRIIPRNPSSDYPSDFEVIQIKOKAGLLTLEDH 120

DB 61 NCYFTAKARNSFTSSALKSEVDNWRRIIPRNPSSDYPSDFEVIQIKOKAGLLTLEDH 120

QY 121 PNKRVTQPKRFRSLKYAESDPTVPCNETRWQSKWSSRLRRLRASLSLGSFHWATGRH 180

DB 121 PNKRVTQPKRFRSLKYAESDPTVPCNETRWQSKWSSRLRRLRASLSLGSFHWATGRH 180

QY 181 SSRLLRAIPROVAOTLQADVLQWQYTGANVRVAFDPTGLSEKHPKFKVKNKERTNWTNE 240

DB 181 SSRLLRAIPROVAOTLQADVLQWQYTGANVRVAFDPTGLSEKHPKFKVKNKERTNWTNE 240

QY 241 RTLDGLGHGTFVAGVIASMRCEGAPDAELHIFRVFNNOVSYTSWFLDAFNAILKK 300

Db 241 RTLDGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
QY 301 IDVLNLSIGGPDPMDFHPPVDKRWELTANNVIMVSAIGNDGLYGTLLNPPADQMDVIGVG 360
Db 301 IDVLNLSIGGPDPMDFHPPVDKRWELTANNVIMVSAIGNDGLYGTLLNPPADQMDVIGVG 360
QY 361 IDPEDNIARSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVGGCRALSGTSVSPVV 420
Db 361 IDPEDNIARSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVGGCRALSGTSVSPVV 420
QY 421 AGAVTLLVSTVQRELNVNPSMKQALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
Db 421 AGAVTLLVSTVQRELNVNPSMKQALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
QY 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNMGVGTGRIVDKPQWYLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNMGVGTGRIVDKPQWYLPQ 540
QY 541 NGDNIEVAFSSYVLWPMWGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
Db 541 NGDNIEVAFSSYVLWPMWGYLAISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
QY 601 QTSVTKLPKVKIIPPPSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHIHTN 660
Db 601 QTSVTKLPKVKIIPPPSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHIHTN 660
QY 661 FRDYOHLRSMGYFVEVLGAPFCFDASQYGTLLMDVSEEEYFPEEIAKLRRDNDGLSL 720
Db 661 FRDYOHLRSMGYFVEVLGAPFCFDASQYGTLLMDVSEEEYFPEEIAKLRRDNDGLSL 720
QY 721 VIFSDWYNTSVMRKVKFYDENTQWMPDPTGGANIPALNELLSSVNMWGFSDGLYEGETL 780
Db 721 VIFSDWYNTSVMRKVKFYDENTQWMPDPTGGANIPALNELLSSVNMWGFSDGLYEGETL 780
QY 781 ANHDMYASGCSIAKPEDEGWITQTKDQGLEVLQKQETAVENVPIGLYQIPABGGGR 840
Db 781 ANHDMYASGCSIAKPEDEGWITQTKDQGLEVLQKQETAVENVPIGLYQIPABGGGR 840
QY 841 IVLYGDSNCLDDSHRQKDFWLLDALLQYTSYGTTPPSLSHSGNRQPPSGAGSVTPERM 900
Db 841 IVLYGDSNCLDDSHRQKDFWLLDALLQYTSYGTTPPSLSHSGNRQPPSGAGSVTPERM 900
QY 901 EGNHLHYRYSKVLGAHLDGPKRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLKV 960
Db 901 EGNHLHYRYSKVLGAHLDGPKRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLKV 960
QY 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVVQ 1020
Db 961 VLPNFRSNRQVRPLSPGESGAWDIPGGIMPGRYNOEVGOTIPVFAFLGAMVVLAFVVQ 1020
QY 1021 INKASRPKRKRPRVRKRPQLMQVHPKPTPSV 1052
Db 1021 INKASRPKRKRPRVRKRPQLMQVHPKPTPSV 1052

RESULT 5

AAU38520

ID AAU38520 standard; Protein; 1052 AA.

XX AC AAU38520;

XX AC AAU38520;

DT 18-DEC-2001 (first entry)

XX Human subtilisin.

DE Human subtilisin.

KW Immunogenic; immunostimulant; immunosuppressive; T-cell epitope;
KW Immune response; hormone; vaccine; cytokine; therapeutic; savinase;
KW hyper-allergenic; hypo-allergenic; cleaning composition; human;
KW textile treatment; contact lens cleaning solution; waste treatment;
KW cosmetic formulation; subtilisin; BPN.
XX Homo sapiens.

PN WO200159130-A2.

XX 16-AUG-2001.

XX 22-JAN-2001; 2001WO-US02204.

XX 08-FEB-2000; 2000US-0500135.

XX (GEMV) GENECOR INT INC.

XX Estell DA, Harding FA;

XX WPI; 2001-607196/69.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

A variant of a polypeptide of interest comprising an altered T-cell epitope to produce a different immunogenic response useful in therapeutics, vaccines, textile treatments and cosmetics -

Example 2; Figure 7; 54pp; English.

The invention relates to a variant of a polypeptide of interest comprising an altered T-cell epitope to produce a different immunogenic response. A method for altering the immunogenicity of a protein especially an enzyme, a hormone, a factor, a vaccine, or cytokine is described. The protein produces an immune response as desired and is useful in therapeutics, vaccines and in forming hyper and hypo-allergenic compounds, e.g. cleaning compositions, textile treatments, contact lens cleaning solutions, waste treatment products and cosmetic formulations. Unlike antibody assays, the method determines immune sensitivity to an epitope prior to any sensitisation taking place. AAU38518-AAU38742 represent the amino acid sequences of Bacillus lentus savinase and human subtilisin (BPN) T-cell epitopes, and related amino acid sequences as described in the method of the invention.

Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 22; Length 1052;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLVLLGKHLGDRLEKKSPEKAPCPGCSHLTLKVFSSVWEYEVAF 60
Db 1 MKLVNWLVLVLLGKHLGDRLEKKSPEKAPCPGCSHLTLKVFSSVWEYEVAF 60
QY 61 NGYFTAKARNSFISSALKSSEVDNWRIPRNNPSSDYPDFEVIQIEKQAGLLTLEDH 120
Db 61 NGYFTAKARNSFISSALKSSEVDNWRIPRNNPSSDYPDFEVIQIEKQAGLLTLEDH 120
QY 121 PNTRKRVTPQRKVFRLKYAESDPTVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH 180
Db 121 PNTRKRVTPQRKVFRLKYAESDPTVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH 180
QY 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVEDTGLSEKHPHFKNVKTWNTE 240
Db 181 SSRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVEDTGLSEKHPHFKNVKTWNTE 240
QY 241 RTLDGGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
Db 241 RTLDGGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVFTNNQVSYSYTWFLDAFNALYALKK 300
QY 301 IDVLNLSIGGPDPMDFHPPVDKRWELTANNVIMVSAIGNDGLYGTLLNPPADQMDVIGVG 360
Db 301 IDVLNLSIGGPDPMDFHPPVDKRWELTANNVIMVSAIGNDGLYGTLLNPPADQMDVIGVG 360
QY 361 IDPEDNIARSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVGGCRALSGTSVSPVV 420
Db 361 IDPEDNIARSSRGMTTWELPGYGRMKPDIVTYGAGVRSYVGGCRALSGTSVSPVV 420
QY 421 AGAVTLLVSTVQRELNVNPSMKQALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
Db 421 AGAVTLLVSTVQRELNVNPSMKQALIASARRLPVGNMFEQGHGKLDLLRAYOILANSYK 480
QY 481 QASLSPSYIDLTECPYMWPCSOPIYYGGMPTVVNTILNMGVGTGRIVDKPQWYLPQ 540

Db 481 QASLSPSYDITLPCYMWPCYQPIYYGMPYVNVTLNGMGVGTGRVDPDQWQYLPQ 540
 Qy 541 NGDNIEVAFSYSSVLPWPGSVGLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Db 541 NGDNIEVAFSYSSVLPWPGSVGLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Qy 601 QTSVTKLPKIKVILPTPPSKRVLDQYHNLRYPPGPPRONLRKMKNDPLDNGDHIHTN 660
 Db 601 QTSVTKLPKIKVILPTPPSKRVLDQYHNLRYPPGPPRONLRKMKNDPLDNGDHIHTN 660
 Qy 661 FEDMYOHLRSMGYFVEVLGAPFTCFDASQYCTLLMVDSESEYFPEEIAKLRDNDVNGLSL 720
 Db 661 FEDMYOHLRSMGYFVEVLGAPFTCFDASQYCTLLMVDSESEYFPEEIAKLRDNDVNGLSL 720
 Qy 721 VIFSDWNTSVMKRVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
 Db 721 VIFSDWNTSVMKRVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
 Qy 781 ANHDMYASGCSIAKPEPDGVVITQTFKQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
 Db 781 ANHDMYASGCSIAKPEPDGVVITQTFKQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
 Qy 841 IVLYGDSNCLDDSHKQKCFMLLDALLOQTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
 Db 841 IVLYGDSNCLDDSHKQKCFMLLDALLOQTSYGVTPPSLSHSGNRQPPSGAGSVTPERM 900
 Qy 901 EGNHLHRYSKVLEAHGLDKPRPLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQKV 960
 Db 901 EGNHLHRYSKVLEAHGLDKPRPLPACPRLSNAKPOPLNETAPSNLWKHOKLLSIDLQKV 960
 Qy 961 VLPNFRNRQVRLPSGEGAWDIPGGIMPGRYNOEGVQTIPTVFAFLGAMVYLAFFVVQ 1020
 Db 961 VLPNFRNRQVRLPSGEGAWDIPGGIMPGRYNOEGVQTIPTVFAFLGAMVYLAFFVVQ 1020
 Qy 1021 INKAKRPRKRRKPRVKRPMQOVHPKPTSV 1052
 Db 1021 INKAKRPRKRRKPRVKRPMQOVHPKPTSV 1052

RESULT 6

AE14527
 ID AE14527 standard; Protein; 1052 AA.
 AC AE14527;
 DT 07-MAY-2002 (first entry)
 DE Human site-1 protease.
 KW Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
 KW sterol regulatory element-binding protein; SREBP; hypercholesterolemia;
 KW dyslipidaemia; atherosclerosis; cardiovascular disease; human.
 XX Homo sapiens.
 OS Homo sapiens.
 PN W0200200873-A1.
 PD 03-JAN-2002.
 PF 19-JUN-2001; 2001WO-SE01386.
 PR 27-JUN-2000; 2000SE-0002417.
 PA (BIOV-) BIOVITRUM AB.
 PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
 DR WPI: 2002-139918/18.
 DR N-PSDB; AAD24182.
 XX Human site-1 protease promoter region for identifying agents capable of
 inhibiting the promoter activity useful in treating medical conditions

PT XX such as obesity, diabetes, atherosclerosis and hypercholesterolemia
 PS Disclosure; Page 22-24; 36pp; English.
 CC The invention relates to human site-1 protease (SLP) promoter region.
 CC The promoter sequence is useful for identification of compounds that
 CC inhibit transcription of SLP which in turn results in inhibition
 CC of sterol regulatory element-binding protein (SREBP) pathway. The
 CC compound identified is useful for the treatment of medical conditions
 CC related to obesity, type II diabetes, hypercholesterolemia,
 CC dyslipidaemia, atherosclerosis and other cardiovascular diseases.
 CC The present sequence is human site-1 protease.

XX Sequence 1052 AA;

Query Match 100.0%; Score 5617; DB 23; Length 1052;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1052; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKLVNIMILLVLLCGKHLGDRLEKSKFEKAPCCPSHLTKVFFSSTVVEYIYAF 60
 Db 1 MKLVNIMILLVLLCGKHLGDRLEKSKFEKAPCCPSHLTKVFFSSTVVEYIYAF 60
 Qy 61 NGYFTAKARNSFISSALKSEYDNRRIIPRNPSSDYPDFEVIQIKKQKAGLLTLEDH 120
 Db 61 NGYFTAKARNSFISSALKSEYDNRRIIPRNPSSDYPDFEVIQIKKQKAGLLTLEDH 120
 Qy 121 PNIRKVTQPKRVFSLKYAESDTPVPCNETRWSOKWSSRPLERASLSLGSFWHATGRH 180
 Db 121 PNIRKVTQPKRVFSLKYAESDTPVPCNETRWSOKWSSRPLERASLSLGSFWHATGRH 180
 Qy 181 SSRLLRAIPROVAQTLQADVLQWQGYTGANRVAVFDTGLSEKHPHFNKVKERTNWTNE 240
 Db 181 SSRLLRAIPROVAQTLQADVLQWQGYTGANRVAVFDTGLSEKHPHFNKVKERTNWTNE 240
 Qy 241 RTLDDGLGHGTFVAGYVATSMRSCQGFAPDAELHIFRVTNNQVSYTSWFLDAFNAILKK 300
 Db 241 RTLDDGLGHGTFVAGYVATSMRSCQGFAPDAELHIFRVTNNQVSYTSWFLDAFNAILKK 300
 Qy 301 IDVLNLSIGGDPDMHDPFVDKWTANNVWISAIGNDGPLYGLTNLPADOMDVIYGGV 360
 Db 301 IDVLNLSIGGDPDMHDPFVDKWTANNVWISAIGNDGPLYGLTNLPADOMDVIYGGV 360
 Qy 361 IDFNITARSSRGMTWELPGYGRMKPDIYTGAGVRGSGVGGCRALSGTSVASPVV 420
 Db 361 IDFNITARSSRGMTWELPGYGRMKPDIYTGAGVRGSGVGGCRALSGTSVASPVV 420
 Qy 421 AGAVTLVSTVQKRELVPASMKQALIASARRLPGVNMPEQGHGKLDLLRAYQILNYSKP 480
 Db 421 AGAVTLVSTVQKRELVPASMKQALIASARRLPGVNMPEQGHGKLDLLRAYQILNYSKP 480
 Qy 481 QASLSPSYDITLPCYMWPCYQPIYYGMPYVNVTLNGMGVGTGRVDPDQWQYLPQ 540
 Db 481 QASLSPSYDITLPCYMWPCYQPIYYGMPYVNVTLNGMGVGTGRVDPDQWQYLPQ 540
 Qy 541 NGDNIEVAFSYSSVLPWPGSVGLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Db 541 NGDNIEVAFSYSSVLPWPGSVGLAISISVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
 Qy 601 QTSVTKLPKIKVILPTPPSKRVLDQYHNLRYPPGPPRONLRKMKNDPLDNGDHIHTN 660
 Db 601 QTSVTKLPKIKVILPTPPSKRVLDQYHNLRYPPGPPRONLRKMKNDPLDNGDHIHTN 660
 Qy 661 FEDMYOHLRSMGYFVEVLGAPFTCFDASQYCTLLMVDSESEYFPEEIAKLRDNDVNGLSL 720
 Db 661 FEDMYOHLRSMGYFVEVLGAPFTCFDASQYCTLLMVDSESEYFPEEIAKLRDNDVNGLSL 720
 Qy 721 VIFSDWNTSVMKRVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
 Db 721 VIFSDWNTSVMKRVFYDENTROWMPDGTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
 Qy 781 ANHDMYASGCSIAKPEPDGVVITQTFKQGLEVLKQETAVVENVPILGLYQIPAEGGGR 840

Db 781 ANHDMYASGCSIAKPEDGVVITQFKDQGLEVLKQETAVENVPIGLYQIYPAEGGGR 840

QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900

Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900

QY 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKY 960

Db 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKY 960

QY 961 VLPNFRSNRPOVRPLSPGESGAWDIPGGIMPGRYNVEVGOTIPVFAFLGAMVVLAFVQ 1020

Db 961 VLPNFRSNRPOVRPLSPGESGAWDIPGGIMPGRYNVEVGOTIPVFAFLGAMVVLAFVQ 1020

QY 1021 INKAKSRPRKRPRVRPQLMQQVHPKTPSV 1052

Db 1021 INKAKSRPRKRPRVRPQLMQQVHPKTPSV 1052

RESULT 7

AY84228

ID AAY84228 standard; peptide; 1052 AA.

XX AC AAY84228;

XX DT 03-JUL-2000 (first entry)

XX DE Amino acid sequence of a hamster site-1 protease.

XX KW Modulator; sterol-regulated Site-1 protease; cholesterol;

XX KW sterol regulatory element binding protein; SREBP; lipid synthesis;

XX KW fatty acid biosynthesis; site-1 protease; protease inhibitor;

XX KW serum cholesterol; hypercholesterolemia; lipid metabolism.

XX OS Cricetus sp.

XX PN W0200009677-A2.

XX PD 24-FEB-2000.

XX PF 13-AUG-1999; 99WO-US18544.

XX PR 14-AUG-1998; 98US-0096571.

XX PR 23-JUL-1999; 99US-0360237.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PA (BROW/) BROWN M S.

XX PA (CHEN/) CHENG D.

XX PA (ESPE/) ESPENSHADE P J.

XX PA (GOLD/) GOLDSTEIN J L.

XX PA (RAWS/) RAWSON R B.

XX PA (SAKA/) SAKAI J.

XX PI Brown MS, Cheng D, Espenshade PJ, Goldstein JL, Rawson RB;

XX PI Sakai J;

XX WP1: 2000-224327/19.

XX PT Novel assay for identifying modulators of sterol-regulated Site-1

XX PT protease useful for the treatment of hypercholesterolemia, involves

XX PT identifying an agent capable of down regulating Site-1 protease

XX PT activity

XX PS Claim 60; Fig 4A-B; 172pp; English.

XX PS

XX CC The specification describes a method for identifying modulators of a

XX CC sterol-regulated Site-1 protease. Site-1 protease cleaves sterol

XX CC regulatory element binding proteins (SREBPs) in the endoplasmic

XX CC reticulum, initiating release from membranes and activating lipid

XX CC synthesis. The modulators therefore also modulate cholesterol and

XX CC fatty acid biosynthesis. The method comprises selecting an agent capable

XX CC of down regulating Site-1 protease and formulating a composition

XX CC comprising the agent. The site-1 protease inhibitors are useful for

CC treating a patient for elevated serum cholesterol. Diseases treated

CC include hypercholesterolemia and other lipid metabolism associated

CC conditions. The present sequence represents a hamster site-1 protease.

XX

SQ Sequence 1052 AA;

Query Match 97.7%; Score 5487; DB 21; Length 1052;

Best Local Similarity 97.1%; Pred. No. 0;

Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY 1 MKLVNWLVLVLLCGKHLGDRLEKSKFEKAPCPGCSHLTLKVFSSSTVVEYIVAF 60

Db 1 MKLVNWLVLVLLCGKHLGDRLEKSKFEKAPCPGCSHLTLKVFSSSTVVEYIVAF 60

QY 61 NGYFTAKARNSFISSALKSSEVDNWRILIPRNNSSDYPDFEVIQIEKQKAGLLTLEDH 120

Db 61 NGYFTAKARNSFISSALKSSEVDNWRILIPRNNSSDYPDFEVIQIEKQKAGLLTLEDH 120

QY 121 PNIRKVTQPKVFRSLKYAESDPTVPCNETRWQKQSSRPLRRASLSLGSFGFWHATGRH 180

Db 121 PNIRKVTQPKVFRSLKYAESDPTVPCNETRWQKQSSRPLRRASLSLGSFGFWHATGRH 180

QY 181 SSRLLRAIPROVAOTLQADVLQMGYTGANVRVAVEDTGLSEKHPHFKNVKTNTWNE 240

Db 181 SSRLLRAIPROVAOTLQADVLQMGYTGANVRVAVEDTGLSEKHPHFKNVKTNTWNE 240

QY 241 RTLDDGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVTNNQSYTSWFLDAFNALKK 300

Db 241 RTLDDGLGHGTFVAGVIASMRCEQGFAPDAELHIFRVTNNQSYTSWFLDAFNALKK 300

QY 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGTLLNAPDQMDVIGVG 360

Db 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGTLLNAPDQMDVIGVG 360

QY 361 IDFEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVSGVKGGRALSGTSVASPVV 420

Db 361 IDFEDNIARSSRGMTTWELPGGYGRMKPDIVTYGAGVSGVKGGRALSGTSVASPVV 420

QY 421 AGAVTLVSTVQKRELNVNPMASMOALIASARRLPGVNMFGHGKLDLLRAYQILNSYKP 480

Db 421 AGAVTLVSTVQKRELNVNPMASMOALIASARRLPGVNMFGHGKLDLLRAYQILNSYKP 480

QY 481 QASLSPSYIDLTECPYMWPCSQPIYYGMPVTVNVTILNGMGVYGRVDPKQWPLPQ 540

Db 481 QASLSPSYIDLTECPYMWPCSQPIYYGMPVTVNVTILNGMGVYGRVDPKQWPLPQ 540

QY 541 NGDNIEVAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTITVASPAETESKNGAE 600

Db 541 NGDNIEVAFSYSSVLPWPSGYLAISVTKKAASWEGIAQGHVMTITVASPAETESKNGAE 600

QY 601 QTSVTKLPKVKIIPTPPRSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHDHTN 660

Db 601 QTSVTKLPKVKIIPTPPRSKRVLDQYHNLRYPPGYFPRDNLRMKNDPLDWDGHDHTN 660

QY 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720

Db 661 FRDMYQHLRSMGYFVEVLGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720

QY 721 VIFSDWNTSVMRKVFYDENTRQWMPDTPGGANIPALNELLSSVWNGFSDGLYEGFTL 780

Db 721 VIFSDWNTSVMRKVFYDENTRQWMPDTPGGANIPALNELLSSVWNGFSDGLYEGFTL 780

QY 781 ANHDMYASGCSIAKPEDGVVITQFKDQGLEVLKQETAVENVPIGLYQIYPAEGGGR 840

Db 781 ANHDMYASGCSIAKPEDGVVITQFKDQGLEVLKQETAVENVPIGLYQIYPAEGGGR 840

QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900

Db 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPPSGAGSVTPERM 900

QY 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKY 960

Db 901 EGNHLHRYSKVLEAHLGDPKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLLSIDLDKY 960

```
QY 961 VLPNRSRNPQVRPLSPGSGAWDIPGIMPGRYNOECORTIPVFAELGAMVLAFFVQ 1020
DB 961 VLPNRSRNPQVRPLSPGSGAWDIPGIMPGRYNOECORTIPVFAELGAMVLAFFVQ 1020
QY 1021 INKASRPRKRPRKVRKPOLMOQVHPKTPSV 1052
DB 1021 ISKASRPRKRPRKVRKPOLMOQVHPKTPSV 1052

RESULT 8
AAE14528
ID AAE14528 standard; Protein; 1052 AA.
XX
AC AAE14528;
XX
DT 07-MAY-2002 (first entry)
XX
DE Hamster site-1 protease.
XX
KW Site-1 protease; S1P; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; hamster.
XX
OS Cricetulus griseus.
XX
PN W0200200873-AA.
XX
PD 03-JAN-2002.
XX
PF 19-JUN-2001; 2001WO-S001386.
XX
PR 27-JUN-2000; 2000SE-0002417.
XX
PA (BIOV-) BIOVITRUM AB.
XX
PI Abrahamson L, Ekblom J, Forsgren M, Hoerling J, Johansson P;
XX WPI; 2002-139918/18.
DR N-PSDB; AAD24185.
XX
XX Human site-1 protease promoter region for identifying agents capable of
PT inhibiting the promoter activity useful in treating medical conditions
PT such as obesity, diabetes, atherosclerosis and hypercholesterolemia -
XX
XX Disclosure; Page 29-32; 36pp; English.
XX
CC The invention relates to human site-1 protease (S1P) promoter region.
CC The promoter sequence is useful for identification of compounds that
CC inhibit transcription of S1P, which in turn results in inhibition
CC of sterol regulatory element-binding protein (SREBP) pathway. The
CC compound identified is useful for the treatment of medical conditions
CC related to obesity, type II diabetes, hypercholesterolaemia,
CC dyslipidaemia, atherosclerosis and other cardiovascular diseases.
CC The present sequence is hamster site-1 protease.
XX
SQ Sequence 1052 AA;
XX
Query Match 97.7%; Score 5487; DB 23; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLVINILLVLLCGKKHLGDRLEKKEKAPCGCSHLTLKVEFSVTVVEYIVAF 60
DB 1 MKLVINILLVLLCGKKHLGDRLEKKEKAPCGCSHLTLKVEFSVTVVEYIVAF 60
QY 61 NGYFTAKARNSFTSALKSSEVDNWRITPRNPSDDPSDFEVIQIKEKAGLLTLEDH 120
DB 61 NGYFTAKARNSFTSALKSSEVDNWRITPRNPSDDPSDFEVIQIKEKAGLLTLEDH 120
QY 121 PNKKVTPQKVRSLKVAESDPTVPCNETRWKQWOSRPLRRASISLSCGFHWATGRH 180
DB 121 PNKKVTPQKVRSLKVAESDPTVPCNETRWKQWOSRPLRRASISLSCGFHWATGRH 180

181 SSRLLRAIPROVAQTLOADYLWOMGYTCANRVAVAFDTGLSEKHHPKFKVNERWTNE 240
181 SSRLLRAIPROVAQTLOADYLWOMGYTCANRVAVAFDTGLSEKHHPKFKVNERWTNE 240
241 RTLDGGLGHGTFVAGVIASMRCCGFADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
241 RTLDGGLGHGTFVAGVIASMRCCGFADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
301 IDVLNLSIGGPDFMDHPFVDKVMELTANNVIMVSAIGNDGLYGTLLNPNADOMDVIYGG 360
301 IDVLNLSIGGPDFMDHPFVDKVMELTANNVIMVSAIGNDGLYGTLLNPNADOMDVIYGG 360
361 IDFDENIARFSSRGMTTWELPGYGRMKPDIIVTYGAGVRGSGYKGCRCALSTSVASPV 420
361 IDFDENIARFSSRGMTTWELPGYGRMKPDIIVTYGAGVRGSGYKGCRCALSTSVASPV 420
421 AGAVTLLVSTVOKRELVPASVKQALIASARRLPGVNMFQGHGKLDLLRAYOILSNYKP 480
421 AGAVTLLVSTVOKRELVPASVKQALIASARRLPGVNMFQGHGKLDLLRAYOILSNYKP 480
481 QASLSPSYIDLTECPYMWPCYQPIYGGMTVTVNVTILNGMGVTCGRIVDKPDQWPLPQ 540
481 QASLSPSYIDLTECPYMWPCYQPIYGGMTVTVNVTILNGMGVTCGRIVDKPDQWPLPQ 540
541 NGDNIEVAFSSVLPWPGSYLAISVTKKAASWEGIAOCHVIMIVASPAETESKNGAE 600
541 NGDNIEVAFSSVLPWPGSYLAISVTKKAASWEGIAOCHVIMIVASPAETESKNGAE 600
601 QTSVTKLPKVKIIPPPRSKRVLDQVHNLRYPGYPFPRDNLRMKNDPLDNGDHIHTN 660
601 HTSVTKLPKVKIIPPPRSKRVLDQVHNLRYPGYPFPRDNLRMKNDPLDNGDHIHTN 660
661 FRDMYQHLRSMGYFEVLGAPTCFDSQYGTLLMVDSEBEYEPPEIAKLRRDNDGLSL 720
661 FRDMYQHLRSMGYFEVLGAPTCFDSQYGTLLMVDSEBEYEPPEIAKLRRDNDGLSL 720
721 VIFSDWYNTSVNRKVFYDENTROWMPDTGGANIPALNELLSSVNMNMGFSDGLYEGEFTL 780
721 VIFSDWYNTSVNRKVFYDENTROWMPDTGGANIPALNELLSSVNMNMGFSDGLYEGEFTL 780
781 ANHDMYASGCSIAKFPEDGVVITQTFKDOGLEVKQETAVENVPILGYQIPAEAGGR 840
781 ANHDMYASGCSIAKFPEDGVVITQTFKDOGLEVKQETAVENVPILGYQIPAEAGGR 840
841 IVLYGDSNCLDSDSHRQKDCFWLLDALQYTSYVTPPSLSHSGNRQRPSPGAGSVTPERM 900
841 IVLYGDSNCLDSDSHRQKDCFWLLDALQYTSYVTPPSLSHSGNRQRPSPGAGSVTPERM 900
901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHOKLLSIDLDKY 960
901 EGNHLHRYSKVLEAHLGDKPRPLPACPRLSWAKPQPLNETAPSNLWKHOKLLSIDLDKY 960
961 VLPNFRSRNPQVRPLSPGSGAWDIPGIMPGRYNQEVGQTIPTVFAELGAMVLAFFVQ 1020
961 VLPNFRSRNPQVRPLSPGSGAWDIPGIMPGRYNQEVGQTIPTVFAELGAMVLAFFVQ 1020
1021 INKASRPRKRPRKVRKPOLMOQVHPKTPSV 1052
1021 ISKASRPRKRPRKVRKPOLMOQVHPKTPSV 1052

RESULT 9
AAE06335
ID AAE06335 standard; Protein; 1052 AA.
XX
AC AAE06335;
XX
DT 03-OCT-2000 (first entry)
XX
DE Rat subtilisin-kexin isoenzyme 1.
XX
KW Rat; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytosolic;
```

vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
 Ras-dependent cancer; restenosis; amyloid protein formation;
 pro-brain-derived neurotrophic factor; proBDNF;
 sterol-regulatory element-binding protein; SREBP.
 Rattus sp.
 WO200026348-A2.
 11-MAY-2000.
 04-NOV-1999; 99WO-CA01058.
 04-NOV-1998; 98CA-2249648.
 (RECL-) INST RECH CLINIQUES MONTREAL.
 Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
 WPI; 2000-365601/31.
 N-PSDB; AAA571198.
 Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
 producing a polypeptide useful for treating hypercholesterolemia, liver
 steatosis and amyloidosis, comprises a specific amino acid sequence -
 Claim 1; Page 61-65; 119pp; English.
 The present sequence is rat subtilisin-kexin isoenzyme 1 (SKI-1),
 a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
 from rat adrenal glands by RT-PCR using active site degenerate
 primers. SKI-1 cleaves at a specific threonine residue within
 the N-terminal segment of pro-brain-derived neurotrophic factor
 (proBDNF). It is also capable of cleaving sterol-regulatory
 element-binding proteins (SREBPs), which function to control lipid
 biosynthesis and uptake in animal cells. Peptides which bind to and are
 cleaved by SKI-1 may be used for monitoring SKI-1 activity, for screening
 inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
 activity. Proteic fragments of SKI-1 which bind to the SKI-1
 catalytic site may be used as inhibitors of SKI-1 activity. They may
 be used to treat diseases involving overexpression of SKI-1 or SKI-1
 substrate. Such diseases include hypercholesterolaemia, high levels of
 fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
 Ras-dependent cancer, restenosis and amyloid protein formation.
 Sequence 1052 AA;
 Query Match 97.5%; Score 5476; DB 21; Length 1052;
 Best Local Similarity 96.7%; Pred. No. 0;
 Matches 1017; Conservative 20; Mismatches 15; Indels 0; Gaps 0;
 1 MKLVNWLWLLVLLCGKHLGDRLEKKSPKAPCGCSHLTKLVFSSVTVEYIVAF 60
 1 MKLVNWLWLLVLLCGKHLGDRLEKKAPKAPCSHLTKLVFSSVTVEYIVAF 60
 61 NGYFTAKARNSFSSALKSEVDNWRRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEDH 120
 61 NGYFTAKARNSFSSALKSEVDNWRRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEDH 120
 121 PNKRYTPQKVRPSLKAEASDTPVPCNETRWSQKWSRPLRRASLSLGSFGWHATGRH 180
 121 PNKRYTPQKVRPSLKAEASDTPVPCNETRWSQKWSRPLRRASLSLGSFGWHATGRH 180
 181 SSRRLRLAIPROVAOTLQADVLQMGYTGANVRVAFDTCLEKHPHFKNVKERTWNTE 240
 181 SSRRLRLAIPROVAOTLQADVLQMGYTGANVRVAFDTCLEKHPHFKNVKERTWNTE 240
 241 RTLDDGLGHGTFVAGVIASMRQCQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
 241 RTLDDGLGHGTFVAGVIASMRQCQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
 301 IDVLNLSIGGPDFMDHFFVDKVMWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360

301 MDVLNLSIGGPDFMDHFFVDKVMWELTANNVIMVSAIGNDGLYGTLLNPNADQMDVIGVG 360
 361 IDFDNLTARSSSGMTTWELPGGYGRMKPDIIVTYGAGVRGSGVKGCRALSGTSVSPVV 420
 361 IDFDNLTARSSSGMTTWELPGGYGRVKPDIIVTYGAGVRGSGVKGCRALSGTSVSPVV 420
 421 AGAVTLIVTVQKRELNVNPMASMQALIASARRLPVNMFGQHGKLDLLRAYQILSYKP 480
 421 AGAVTLIVTVQKRELNVNPMASMQALIASARRLPVNMFGQHGKLDLLRAYQILSYKP 480
 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVVNVTILNMGVGTGRIVDKPWQYLPQ 540
 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMPTVVNVTILNMGVGTGRIVDKPWQYLPQ 540
 541 NGDNIEVAFSYSSVLNPMWPGYLAISISVTKKAASWEGIAOGHVMITVASPAETESKNGAE 600
 541 NGDNIEVAFSYSSVLNPMWPGYLAISISVTKKAASWEGIAOGHVMITVASPAETESKNGAE 600
 601 QTSTVKLPIKVKIIPTPPSKRVLMDQYHNLRYPPGYFPRDNLNRMKNDPLDWDNGDHHTN 660
 601 HTSTVKLPIKVKIIPTPPSKRVLMDQYHNLRYPPGYFPRDNLNRMKNDPLDWDNGDHHTN 660
 661 FRMYQHLSRMGVEVILGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
 661 FRMYQHLSRMGVEVILGAPFTCFDASQYGTLLMVDSEEEYFPEETAKLRDNDGLSL 720
 721 VIFSDWYNTSVMRKVKFYDENTQWMPDTGGANIPALNELLSSVNMNFGSDGLYEGEFTL 780
 721 VIFSDWYNTSVMRKVKFYDENTQWMPDTGGANIPALNELLSSVNMNFGSDGLYEGEFTL 780
 781 ANHDMYASGCSIAKFPEDGVVITQTFKDGLEVLKQETAVENVPIILGLYQIPAEGGGR 840
 781 ANHDMYASGCSIAKFPEDGVVITQTFKDGLEVLKQETAVENVPIILGLYQIPAEGGGR 840
 841 IVLYGSDNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTERM 900
 841 IVLYGSDNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRQRPSPGAGSVTERM 900
 901 EGNHLHRSKVLBAHLGDPKPRLPACPRLSWAKPQPLNETAPSNLWKHKLISIDLDKV 960
 901 EGNHLHRSKVLBAHLGDPKPRLPACPRLSWAKPQPLNETAPSNLWKHKLISIDLDKV 960
 961 VLPNFRSNRPQVRPLSPGSGAWNDIPGGIMPGRYNQEVGQTPVFAFLGAMVALAFFVQ 1020
 961 VLPNFRSNRPQVRPLSPGSGAWNDIPGGIMPGRYNQEVGQTPVFAFLGAMVALAFFVQ 1020
 1021 INKASRPKRKRPRKRPQMLQOQVHPKTPSV 1052
 1021 INKASRPKRKRPRKRPQMLQOQVHPKTPSV 1052
 1021 ISKASRPKRKRPRKRPQMLQOQVHPKTPSV 1052
 1021 ISKASRPKRKRPRKRPQMLQOQVHPKTPSV 1052
 RESULT 10
 AAB06336
 ID AAB06336 standard; Protein; 1052 AA.
 xx AAB06336;
 xx AC AC
 xx 03-OCT-2000 (first entry)
 xx Mouse subtilisin-kexin isoenzyme 1.
 xx Mouse; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytostatic;
 xx vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
 xx Ras-dependent cancer; restenosis; amyloid protein formation;
 xx pro-brain-derived neurotrophic factor; proBDNF;
 xx sterol-regulatory element-binding protein; SREBP.
 xx Mus sp.
 xx WO200026348-A2.
 xx 11-MAY-2000.
 xx

PF 04-NOV-1999; 99WO-CA01058.
XX
PR 04-NOV-1998; 98CA-2249648.
XX
PA (RECL-) INST RECH CLINIQUES MONTREAL.
XX
PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX
DR WPI; 2000-365601/31.
XX
DR N-PSDB; AAA57199.
XX
XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
PT producing a polypeptide useful for treating hypercholesterolemia, liver
PT steatosis and amyloidosis, comprises a specific amino acid sequence -
XX
PS Claim 1; Page 67-71; 119pp; English.
XX
XX The present sequence is mouse subtilisin-kexin isoenzyme 1 (SKI-1),
CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
CC from mouse corticotropic cells by RT-PCR using active site degenerate
CC primers. SKI-1 cleaves at a specific threonine residue within
CC the N-terminal segment of pro-brain-derived neurotrophic factor
CC (proBDNF). It is also capable of cleaving sterol-regulatory
CC element-binding proteins (SREBPs), which function to control lipid
CC biosynthesis and uptake in animal cells. Peptides which bind to and are
CC cleaved by SKI-1 may be used for monitoring SKI-1 activity for screening
CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
CC activity. Proteic fragments of SKI-1 which bind to the SKI-1
CC catalytic site may be used as inhibitors of SKI-1 activity. They may
CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
CC substrate. Such diseases include hypercholesterolemia, high levels of
CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
CC Ras-dependent cancer, restenosis and amyloid protein formation.
XX
SQ Sequence 1052 AA;

Query Match 96.6%; Score 5427; DB 21; Length 1052;
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1008; Conservative 25; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKLVNIWLLVLLCGKKHLDLEKFKAPCPGCSHLTKVFEFSSTVVEYIVAF 60
DB 1 MKLVSTWLLVLLCGKKHLDLGRTRALEKAPCPSCSHLTAVFSSVVEYIVAF 60
QY 61 NGYFTAKARNSFSSALKSEVDNRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEH 120
DB 61 NGYFTAKARNSFSSALKSEVDNRIIPRNPSSDYPSPDFEVIQKEKAGLLTLEH 120
QY 121 PNKEVTPQKVFSLKYAESDTPVPCNETRMSOKWSSRPLRSLSGSGFWHATGRH 180
DB 121 PNKAVTPQKVFSLKYAESDTPVPCNETRMSOKWSSRPLRSLSGSGFWHATGRH 180
QY 181 SSRLLRAIPQVAQTQADVLWQMGYTGANRVAVDFGLSEKHPFKNVKERTNWTNE 240
DB 181 SSRLLRAIPQVAQTQADVLWQMGYTGANRVAVDFGLSEKHPFKNVKERTNWTNE 240
QY 241 RTLDGLGHGTGVAGVIASMRGCGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
DB 241 RTLDGLGHGTGVAGVIASMRGCGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGDFMDHPFVDFKVMELTANNVIMVSAIGNDGPLYGLTNPNADQMDVIGVG 360
DB 301 MDVLNLSIGGDFMDHPFVDFKVMELTANNVIMVSAIGNDGPLYGLTNPNADQMDVIGVG 360
QY 361 IDFDENTARFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRSGVKGCGRALSGTSSVAPV 420
DB 361 IDFDENTARFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRSGVKGCGRALSGTSSVAPV 420
QY 421 AGAVTLVSTVQKRELNVNPSAKMKAALIASARLLPGVNMFEQHGKLDLLRAYOILNSYKP 480
DB 421 AGAVTLVSTVQKRELNVNPSAKMKAALIASARLLPGVNMFEQHGKLDLLRAYOILNSYKP 480
QY 481 QASLSFSYIDLTECPYMWPCYQPIYGGMPTVNVVILAGMGVTRIVYDKENPYPQP 540

DB 481 QASLSFSYIDLTECPYMWPCYQPIYGGMPTVNVVILAGMGVTRIVYDKENPYPQP 540
QY 541 NGDNTEVAFSSYSSVLPWWSGYLAISISVTTKAAASWEGIAOCHVMITVAPPAETESKNAE 600
DB 541 NGDNTEVAFSSYSSVLPWWSGYLAISISVTTKAAASWEGIAOCHVMITVAPPAETESKNAE 600
QY 601 QTSYKLPKIKYIIPTTPPSKRVLMDOYHNLRYPGYPFRDNRNMRKNDPLDNGSDHIHTN 660
DB 601 HTSYKLPKIKYIIPTTPPSKRVLMDOYHNLRYPGYPFRDNRNMRKNDPLDNGSDHIHTN 660
QY 661 FRDYMQLHRSMGYFEVVLGAPFTCFDASQYGTLLMVDSEEEYFPEETIAKLRRDVGDLGL 720
DB 661 FRDYMQLHRSMGYFEVVLGAPFTCFDASQYGTLLMVDSEEEYFPEETIAKLRRDVGDLGL 720
QY 721 VIFSDWNTSYVARKVKFYDENTROMMPDTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
DB 721 VIFSDWNTSYVARKVKFYDENTROMMPDTGGANIPALNELLSSVWNGFSDGLYGEFTL 780
QY 781 ANHDMYASGCSIAKPPEDGVVITOTFDOGLEVLKQETAVENVPIGLYQIPAEQGR 840
DB 781 ANHDMYASGCSIAKPPEDGVVITOTFDOGLEVLKQETAVENVPIGLYQIPAEQGR 840
QY 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPESGAGSVTPERM 900
DB 841 IVLYGDSNCLDDSHRQKDCFWLLDALLQYTSYGVTPPSLSHSGNRORPESGAGSVTPERM 900
QY 901 EGNHLRYSKVLEAHLGDPKPRPLPACPLSWAKPQPLNETAFSNLWKHKLLSIDDKV 960
DB 901 EGNHLRYSKVLEAHLGDPKPRPLPACPLSWAKPQPLNETAFSNLWKHKLLSIDDKV 960
QY 961 VLPNFRNRQVPRPLSPGSGAWDIPGGIMPRGNQEVGOTIPVFAFLGAMVYVAFVYQ 1020
DB 961 VLPNFRNRQVPRPLSPGSGAWDIPGGIMPRGNQEVGOTIPVFAFLGAMVYVAFVYQ 1020
QY 1021 INKAKSRPKRKRPRKPRQLMQQVHPKPTPSV 1052
DB 1021 ISKAKSRPKRKRPRKPRQLMQQVHPKPTPSV 1052

RESULT 11
ABB90255 standard; Protein; 666 AA.
XX ABB90255;
XX AC ABB90255;
XX DT 24-MAY-2002 (first entry)
XX DE Human polypeptide SEQ ID NO 2631.
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW allergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
XX WO200190304-A2.
XX
XX 29-NOV-2001.
XX
XX 18-MAY-2001; 2001WO-US16450.
XX
XX 19-MAY-2000; 2000US-205515P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Birse CE, Rosen CA;
XX
XX WPI; 2002-122018/16.
XX
XX N-PSDB; ABL90664.

Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -

Claim 11: SEQ ID NO 2631: 2081pp + Sequence Listing: English.

The invention relates to novel genes (ABL9449-ABL90853) and proteins (ABB9040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune chryoiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 666 AA;

Query Match 63.2%; Score 3549; DB 23; Length 666;
Best Local Similarity 99.5%; Pred. No. 1.4e-254;
Matches 663; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY	387	MKPDIVYTGAGVGRSGKCRALSGTSVASPVVAGAVTLVSTVKRELNVNASKQAL	446
Db	1	MKPDIVYTGAGVGRSGKCRALSGTSVASPVVAGAVTLVSTVKRELNVNASKQAL	60
QY	447	IASARLPGVNMFEQGHGKLDLLRAYQILNSYKQASLSPSYIDLTECPYMWPCSQPIY	506
Db	61	IASARLPGVNMFEQGHGKLDLLRAYQILNSYKQASLSPSYIDLTECPYMWPCSQPIY	120
QY	507	YCGMPVTVVNTILNGHGVGRIVYDKPDWQPYLPQNGDNTEVAFSSVLPWPSGYLAISI	566
Db	121	YCGMPVTVVNTILNGHGVGRIVYDKPDWQPYLPQNGDNTEVAFSSVLPWPSGYLAISI	180
QY	567	SVTKKAASWEGIAQGHVMTIVASPAETESKNGAEQSTVKLPKVIPIPTPPSKRVLWD	626
Db	181	SVTKKAASWEGIAQGHVMTIVASPAETESKNGAEQSTVKLPKVIPIPTPPSKRVLWD	240
QY	627	QYHNLRYPPGYFRDRLRMKNPDLWNGDHIHTNFRDMYQHLRSMGYFVEVLGAPFTCFD	686
Db	241	QYHNLRYPPGYFRDRLRMKNPDLWNGDHIHTNFRDMYQHLRSMGYFVEVLGAPFTCFD	300
QY	687	ASQYGTLLMWDSSEEFPEEIAKLRRDNDGLSLVIFSDWNTSVARKVKFYDENTROWW	746
Db	301	ASQYGTLLMWDSSEEFPEEIAKLRRDNDGLSLVIFSDWNTSVARKVKFYDENTROWW	360
QY	747	MPDTGGANTPALNELLSSVNMWGSGLYEGEFTLANHDMYASGCSIAKFPDGVVITQT	806
Db	361	MPDTGGANTPALNELLSSVNMWGSGLYEGEFTLANHDMYASGCSIAKFPDGVVITQT	420
QY	807	FKDQGLEVLKQETAVVENVPIGLYQIPAEAGGGRIVLYGDSNCLDSDHQRKQCFWLLDAL	866
Db	421	FKDQGLEVLKQETAVVENVPIGLYQIPAEAGGGRIVLYGDSNCLDSDHQRKQCFWLLDAL	480
QY	867	LQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPPERMEGNHLHRYSKVLEAHLGDKPRLPA	926
Db	481	LQYTSYGVTPPSLSHSGNRQRPSPGAGSVTPPERMEGNHLHRYSKVLEAHLGDKPRLPA	540
QY	927	CPRLSWAKPOPLNETAPSNLWKHKLLSIDLDRVLPNFRSNRPQVRPLSPGSGGAWDIP	986
Db	541	CPRLSWAKPOPLNETAPSNLWKHKLLSIDLDRVLPNFRSNRPQVRPLSPGSGGAWDIP	600
QY	987	GGIMPGRYNQEVEGQTIPTVPAFLGAMVVLAFVQVQINKAKSRPRKRRKPRVLRPQLMQQVHP	1046

Db	601	GGIMPGRYNQEVEGQTIPTVPAFLGAMVVLAFVQVQINKAKSRPRKRRKPRVLRPQLMQQVHP	660
QY	1047	PKTPSV 1052	
Db	661	PKTPSV 666	

RESULT 12
ARB20015
ID AAB20015 standard; Protein; 992 AA.
XX AAB20015;
AC AAB20015;
DT 28-MAR-2001 (first entry)
XX Drosophila site 1 protease SIP homologue (dSIP).
DE Fly; site 1 protease; SIP; dSIP; SREBP;
XX sterol regulatory element binding protein; transgenic animal;
KW animal model; lipid metabolism; transcription factor; cholesterol;
KW obesity; insulin resistance; therapy; diagnosis; pesticide.
XX Drosophila melanogaster.
OS
XX Key Location/Qualifiers
XX Domain 1..21
FT /note= "transmembrane domain"
FT Domain 363..383
FT /note= "transmembrane domain"
FT Domain 458..478
FT /note= "transmembrane domain"
FT Domain 500..520
FT /note= "transmembrane domain"
FT Region 154..425
FT /note= "predicted homology to peptidase S8
FT subtilase family"
XX W0200076308-A1.
XX 21-DEC-2000.
XX 08-JUN-2000; 2000WO-US15880.
XX 14-JUN-1999; 99US-0332522.
XX 15-MAR-2000; 2000US-0189700.
XX (EXEL-) EXELIXIS INC.
XX Costa MA, Doberstein SK, Elson S, Ferguson KC, Homburger SA;
PI Ebens AJ, Keegan KP, Stout TJ;
XX WPI; 2001-091292/10.
XX N-PSDB; AAA89290.
XX Novel invertebrate organism genetically modified to express or
PT mis-express steroid regulatory element binding protein pathway protein
PT used as model system for studying lipid metabolism and determining
PT lipid content -
XX Claim 8; Page 80-84; 90pp; English.
XX The present sequence is that of Drosophila melanogaster dSIP, a
CC homologue of mammalian site 1 proteases (SIP) that are involved in
CC sterol regulatory element binding protein (SREBP) processing,
CC cleaving SREBP at the luminal loop. SREBPs are transcription
CC factors that activate genes involved in cholesterol and fatty acid
CC synthesis, and are major mediators of insulin action in the liver.
CC Alterations in SREBP function and expression are implicated in
CC obesity and insulin resistance. The dSIP DNA was identified from
CC an expressed sequence tag database search on the basis of homology
CC to hamster S2P. The invention provides fly and nematode
CC (Caenorhabditis elegans) polynucleotides (see AAA89287-90) and

CC encoded proteins (see AAB20012-15) involved in the SREBP pathway.
CC C. elegans and D. melanogaster animals genetically modified to
CC express or mis-express these proteins are claimed. These
CC genetically modified animal models have identifiable phenotypes
CC that make them useful in assays for studying lipid metabolism,
CC other genes implicated in lipid metabolism and compounds capable
CC of modulating lipid biosynthetic pathways. Model organisms or
CC cultured cells can be used to identify new drug targets,
CC therapeutic agents, diagnostics and prognostics of disorders
CC associated with lipid metabolism, and also to identify pesticide
CC targets directed to components of the SREBP pathway.
XX
SQ

Sequence 992 AA;

Query Match 42.1%; Score 2366.5; DB 22; Length 992;
Best Local Similarity 46.6%; Pred. No. 1.4e-166;
Matches 466; Conservative 151; Mismatches 232; Indels 151; Gaps 16;
Qy 47 FSSVVEVEIVAFNGYFTAKARNSFTSSALKSSVDNWRILIPNPSDDSPDFEVIQI 106
Db 19 FKTAVVNEFIVHFHFKYFAPVRESYIAAKLLGSNTWNRIVPRNLAWQYPSDFDILRV 78
Qy 107 KEQKAG----LLTLEDPNKRVTPOKRVRSLYK-RESPTVPCNETHWQKQSSRP 161
Db 79 CDGESSESFEIERQTHPSKAVVQPSVRRLINDAYSMLT----- 121
Qy 162 LARASLSGSGFWHATGRSSRRLLRAIP-----RVAQTLOADVLMQMGYTCANRVAV 216
Db 122 -----YTH--RHQGVLRNPNNDHRQLC-SVLHANILWKLGITGKGVKAI 167
Qy 217 FDTGLSEKHPFNKERTNWTNERTLDDGLGHGTFTAGVIASMRCCGFAPDAELHIFR 276
Db 168 FDTGLTKNHPFNKERTNWTNEKSLDDRVSHGTFTAGVIASSRCLGAFAPDAELYFK 227
Qy 277 VFTNQSVYTSWFLDAFNAYLKKIDVLNLSIGGPDFMDHPFDVKWELTANNLWVSAI 336
Db 228 VFTNSQSVYTSWFLDAFNAYLKYKINILNLSIGGPDFMDSPFEKVLSEANNVIMSA 287
Qy 337 GNDGLYGLTLNPNADQMDVIGVGIGDIFEDNIAFSSRGMTTWELPGYGRMKPDVTYGA 396
Db 288 GNDGLYGLTLNPGDSQDVGVGGIGQFDKIAKFSRGMTTWELPLGYGRMGDLIVTGS 347
Qy 397 GVRSGVKGCRALSGTSVSPVAVGAVTLVYS-TVQKRELVNPSAKKALIASARRLP 455
Db 348 QVEGSDVKGCRRLSGTSVSPVAVGAAALLISGAFQKIDYINPASKLVLEGAKLPH 407
Qy 456 VNMFEQGHGKLDLLRAYQILNSYKQASLSPSYIDLTCTPWWPYCSQPIYYGMPVTYN 515
Db 408 YNMFQAGKLLKSMQLLSYKPKLILPAILDFTQ-NYMWPFSSQPIYYGSSVAIAN 466
Qy 516 VTILNGMVGTRIVDKPQPYLPQNGDNIEVAFYSVLPNPSGYLAISISVTKKAASW 575
Db 467 VTILNGISVTSHTVIGPKWIPDFENQGLQVSAQVSPVWPMTGMSVFIKVEGENF 526
Qy 576 EGIAQGHVMTVASPAETESKNGAEOQTSVKLPFKVKIIPTPPRSKRVLMQDQYHNLRYPP 635
Db 527 EGVCCKGSITLVLESFKQTNET---HVTEDVPLIKVTPKPPNKRILMDQYHSLRYPP 583
Qy 636 GVFPDRNLKMDPLDNGDHIHTNFRDMYQHLRSMGVFEVLGAPTCFDSASQGTLLM 695
Db 584 RYIPRDLKVKLPLDWRADHIHTNFRDMYTHLRNVGYIIDVLEPFTCNASDYGALLI 643
Qy 696 VDSSEEFPEETAKLRDY-DNGLSIVFSDMYNTSVMRKVKFYDENTQWMPDPTGGAN 754
Db 644 VDPERGGEDEINALQENYVRGLNVVFGDWMYNTVMKKIKFDENTQWMPDPTGGAN 703
Qy 755 IPALNELLSVWNGFSDGLYEGETLANHDMYYAGSCSIKAFPED--GVVITQTFKDGGL 812
Db 704 IPALNLLKRFPGAFGDFVGEHFKLDHSMYYASGATIVKFPNPNFGDIIVGTLKNDGGL 763
Qy 813 EVLKQET---AVVENVPILGLYQIPA----- 835
Db 764 SIINKTSPSKAKLDVPIFGMFQTKANSIQSNEEIVVNAESNLAEALPTDYSTFKNRVLL 823

Qy 836 -----EGGRIVLYGDSNCLDSDSHROKDCFWLIDALLOYTSYGVTPP 877
Db 824 LTKORSISPAKSNNIETNKGRIAYYGDSCNLDSTHLEKACWLLITFLDFAIN----- 878
Qy 878 SLUSHGNNRQPPSGAGSVTPERMENHLYRYSKYVLEAHLGDPKPRPLPA----- 926
Db 879 --SH-----KSSLQNI-NRTEPHKLERAPLPLRISOSITKRSQD 917
Qy 927 --CPLRSWAKPOLNPTAPSNLWKHQKLLSIDLDKVVLPN 964
Db 918 NNCEQFWLAPTKQNN-----EERKSSIIDVTILEN 949

RESULT 13

ABB63047

ID ABB63047 standard; Protein; 952 AA.

XX ABB63047;

XX AC ABB63047;

XX XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 15933.

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

KW Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

DR WPI: 2001-656860/75.

DR N-PSDB; ABL07150.

XX New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

PT interactions -

XX Disclosure; SEQ ID NO 15933; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent

CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and

CC cell-cell interactions in higher eukaryotes for the development of

CC insecticides, therapeutics and pharmaceutical drugs. The invention

CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

CC sequences (ABL01840-ABL16175) and the encoded proteins

CC (ABBS7737-ABBS72072).

CC The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 952 AA;

Query Match 39.7%; Score 2231; DB 22; Length 952;

Best Local Similarity 44.5%; Pred. No. 1.5e-156;

Matches 445; Conservative 147; Mismatches 230; Indels 178; Gaps 17;

Qy 47 FSSVVEVEIVAFNGYFTAKARNSFTSSALKSSVDNWRILIPNPSDDSPDFEVIQI 106

Db 19 FKTAVVNEFIVHFHFKYFAPVRESYIAAKLLGSNTWNRIVPRNLAWQYPSDFDILRV 78

Query Match	6.1%	Score 340.5	DB 9	Length 381
Best Local Similarity	27.1%	Pred. No. 8.1e-17		
Matches 118	Conservative 65	Mismatches 156	Indels 97	Gaps 17

QY	48	SSTVVEYEVIAENGFTAKARNSFTSSALKSSEVDNWRRIIPRNPSSDYPSEFVIOIK	107
DB	33	SST--EKKYIVGF-----KQTSAMSSAKKQDVI-----	59
QY	108	EKQKAGLLTLE-DHPNIKRVTPQKRVSRSLKAESDPTVPCNETRWQKQSSRPLRRAS	166
DB	60	-SEKGGKVQKQFYVNAATAATLDEKAVKELK--KQPSVA-----	95
QY	167	LSLGSGFNIATGRHSRRLLRALPRQVAQTLQADVLWOMGYTGANRVAVFDTLGLSEKHP	226
DB	96	-----YVEDHIAHEVAQSPYVGISQ-IPALHLSQGYTGSNVKVAVIDSGIDSSHP	146
QY	227	HPKNVKERTNWTNERT--LDDGLGHGTFVAGVTASMRP---CQGFADAEHLHIFRVFNN	281
DB	147	DL-NVRGGASFVPEPNPDQGGSSGHGTHVAGTIAALNNSIGVLGVAPSPASLYAVKVLDS	205
QY	282	QVSYTSWFLDAFNAILKKTIDVLNLSIGGPDFMD--HPFVDKVMELTANNVIMVSAIGND	339
DB	206	GSQYQSWIINGIEWAINNMDVINMSLGGPTGSTALKTVDKA--VSSGIVVAAAAGNE	262
QY	340	GP--LYGTLNPNADQMDVIGVGIDEDNTARFSRSGTWTWELPGGYGRMKPDIVTYCAG	397
DB	263	GSQGSTSTGYCAPYPTSTIAVGVNSNQRASFSSAGS---EL-----DYMAGVGS	310
QY	398	VRGSGVKGGRALSGTSPASPVVAGAVTLVLVSTVQKRELVNPASMKOALIASARRLPCVN	457
DB	311	IQSTLPGGYGANGYISMATPHVAGAAALILS---KHPWTINQVRDLRESTAYILG--N	365
QY	458	MFEQGHGKLDLLRAYQ 473	
DB	366	SFYIGKGLINVQAAAQ 381	

RESULT 17

AAR07970	AAAR07970 standard; protein; 381 AA.
XX	AAAR07970;
AC	
AC	
DT	18-FEB-1991 (first entry)
XX	
DE	Subtilisin gene product.
XX	
KW	Detergents; depilatory tanning; serology.
XX	
OS	Bacillus subtilis.
XX	
FF	Key Location/Qualifiers
FT	Misc-difference 189..190
FT	label= Val-Val
FT	note= "Stahl et al. J.Bacteriol..158,411-418 (1984)"
FT	Protein 107..381
FT	label= Mature subtilisin
XX	
EP	EP398539-A.
PN	
XX	
PD	22-NOV-1990.
XX	
PF	01-MAY-1990; 90EP-0304715.
XX	
PR	17-MAY-1989; 89US-0353124.
XX	
PA	(AMGE-) AMGEN INC.
XX	
PI	Zukowski MM, Narhi LO, Levitt M;
XX	
DR	WPI; 1990-350298/47.
DR	N-PSDB; AAO06587.

[illegible]

FT Misc-difference 265 /note= "Ser or Leu"
XX US5719021-A.
XX 17-FEB-1998.
XX 31-JUL-1992; 92US-0923260.
XX 02-MAY-1989; 89US-0346552.
XX (UYNE-) UNIV NEW JERSEY.
XX Inouye M;
XX WPI; 1998-158792/14.
XX Activation of polypeptides - by interaction with activating peptide,
PT resulting in refolding of the polypeptides to give active form
XX Disclosure; Columns 39-42; 29pp; English.
XX The present sequence represents pro-subtilisin of Bacillus
CC subtilis, and was used to demonstrate the method of the invention.
CC The propeptide is essential for the production of active, correctly
CC folded subtilisin. The M18T substitution resulted in a suppression
CC of this activity. An in vitro method to restore or increase the
CC natural biological activity of a target polypeptide (inactive or with
CC decreased activity due to improper folding), which is normally expressed
CC containing a prosequence forms the basis of the invention. An exogenous
CC activating peptide used to promote refolding of the target polypeptide
CC to give its active form. The activating peptide comprises the prosequence
CC of the target or other proteins with a similar sequence and function to
CC the target polypeptide. The method is used to produce biologically,
CC correctly folded proteins from their inactive, incorrectly folded forms.
CC Suitable target polypeptides include members of the serine protease or
CC subtilisin families.
XX Sequence 352 AA;
Query Match 6.0%; Score 338.5; DB 19; Length 352;
Best Local Similarity 26.8%; Pred. No. 1e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVVEYIVAFNGYFTAKARNSFISALKSEVDNWRIRPNRNPSSDYPSPDFEVIQIK 107
DB 4 SST--EKKIYVGF-----KQTSAMSSAKKKDVI----- 30
QY 108 EKQAGLITLE-DHPNKKRVTPQRKFRSLKYAESDPTVPCNETRWQKQWSSRPLRRAS 166
DB 31 -SEKGGKVKQKRYVNAATAATLDEKAVKELK---KDPVA----- 66
QY 167 LSLGSGFWHATGRHSRRLLRAIPROVAOTLQADVLQWGYTGANRVAVDFDGLSEKHP 226
DB 67 -----YVEDHIAHEYAQSVPYGISQ-IPALHSGQYTGSNVAVIDSGIDSSHP 117
QY 227 HFKNKERTWNTERT--LDGLGHGTFFVAGVIASWRE---CQGFADPDLHIFRVFTNN 281
DB 118 DL-NVRGASGFVSETNPNYQDQSGSHGTHVAGTIALNNSIGVLGVSFASLAVKVLDS 176
QY 282 QVSYTSWFLDAFNALYKIDVNLISIGGPDFMD--HPFDKRWELTANNVIMVSAIGND 339
DB 177 GSGQYSLIINGIEWALSNNMDVINNSLGGPTGTSTALKTVVDKA---VSSGIYVAAAGNE 233
QY 340 GP--LYGTLLNPADQMDVIGGGIDFEDNIAREFSRGMGTWELPGGYGRMKPDIVTYGAG 397
DB 234 GSSGSTSTGVYPAKYPSTIAGVAVNSNQRAFSSAGS-----EL-----DVMAQVGS 281
QY 398 VRGSGVGGCRALSGTFSVASPVWAGAVTLLVSTVQKRELNVNPSMKOALIASARRLPGVN 457
DB 282 IQSTLPGGTVGAVNGTSMATPHVAGAAALILS---KHPTWTNAQVRDLRLESTATVILG--N 336
QY 458 MFEQHGCKLDDLRLAYQ 473

DB 337 SFYKGLINVQAAQ 352
RESULT 19
AAW46595
ID AAW46595 standard; protein; 352 AA.
XX AAW46595;
XX 22-MAY-1998 (first entry)
XX Amino acid sequence of pro-subtilisin E from Bacillus subtilis.
XX Pro-subtilisin; propeptide; subtilisin; subtilisin E; active;
KW folding; increase; activity; improper folding; prosequence;
KW activating peptide; serine protease.
XX Bacillus subtilis.
XX Key Location/Qualifiers
FT Region 1..77 /note= "propeptide"
XX US5719021-A.
XX 17-FEB-1998.
XX 31-JUL-1992; 92US-0923260.
XX 02-MAY-1989; 89US-0346552.
XX (UYNE-) UNIV NEW JERSEY.
XX Inouye M;
XX WPI; 1998-158792/14.
XX Activation of polypeptides - by interaction with activating peptide,
PT resulting in refolding of the polypeptides to give active form
XX Disclosure; Columns 25-28; 29pp; English.
XX The present sequence represents pro-subtilisin of Bacillus subtilis
CC strain 168, and was used to demonstrate the method of the invention. To
CC investigate the role of the propeptide of subtilisin, an expression
CC system for the active subtilisin E was established in Escherichia coli
CC by replacing the pro-sequence of pre-pro-subtilisin with the E. coli
CC OmpA signal peptide. When the amino terminal 14 residues of the
CC propeptide are deleted, no active subtilisin forms, and the shortened
CC propeptides are not removed. The propeptide is essential for the
CC production of active, correctly folded subtilisin. An in vitro method
CC to restore or increase the natural biological activity of a target
CC polypeptide (inactive or with decreased activity due to improper
CC folding), which is normally expressed containing a prosequence forms the
CC basis of the invention. An exogenous activating peptide used to promote
CC refolding of the target polypeptide to give its active form. The
CC activating peptide comprises the prosequence of the target or other
CC proteins with a similar sequence and function to the target polypeptide.
CC The method is used to produce biologically, correctly folded proteins
CC from their inactive, incorrectly folded forms. Suitable target
CC polypeptides include members of the serine protease or subtilisin
CC families.
XX Sequence 352 AA;
Query Match 6.0%; Score 337.5; DB 19; Length 352;
Best Local Similarity 26.8%; Pred. No. 1.2e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSTVVEYIVAFNGYFTAKARNSFISALKSEVDNWRIRPNRNPSSDYPSPDFEVIQIK 107
DB 4 SST--EKKIYVGF-----KQTSAMSSAKKKDVI----- 30

```
QY 108 EKQKAGLLTLE-DHPNLIKRVTPQKVRSLKYAESDTPVCNETRWSQKWSRPLRRAS 166
Db 31 -SEKGGKVKQKQFYVNAATAATLDEKAVKELK---KDPFVA----- 66
QY 167 LSLGSGFWHATGRHSSRRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHP 226
Db 67 -----YVEEDHIAHEYAQSPYGISQ-IPALHSGQYTGTSNKKVAVIDSGIDSSHP 117
QY 227 HFKNVKTNTWNTERT--LDDGLGHGTFVAGVIASMR---COGFAPDAELHIFRVTNN 281
Db 118 DL-NVRGGSFVPESETNPYQDGGSHGTHVAGTIAALNNSICVLGVSFASLIYAVKVLDT 176
QY 282 QVSYTSWFLDAFNAILKIDVLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAICND 339
Db 177 GSGQYSWIINGIEWAINNNMDVINNSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNE 233
QY 340 GP--LYGTLPNADQMDVIGVGIDFEDNTARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 234 GSGGSTSTVGYPAPKYPSTIAVAVNSNQRAFSSAGS---EL-----DYMAGVS 281
QY 398 VRGSGVKGCRALSGTSVAGPVVAVAGVTLVSTVQKRELVPASMKQALIASARRLPGVN 457
Db 282 IQSTLPCTGTCAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDLRESTATYIG--N 336
QY 458 MFEQGHGKLDLLRAYQ 473
Db 337 SFYVGKGLINVQAAAAQ 352

RESULT 20
AAP71060
ID AAP71060 standard; Protein; 381 AA.
XX AC AAP71060;
XX DT 02-MAY-1991 (first entry)
XX DE Wild-type subtilisin from Bacillus subtilis.
XX KW subtilisin; extracellular protease; amylase production.
XX OS Bacillus subtilis.
XX Key Location/Qualifiers
FH Peptide 1..23
FT Peptide /label= signal peptide
FT Peptide 24..106
FT Protein /label= pro sequence
FT Protein 107..381
FT Protein /label= mature subtilisin
XX EP246678-A.
XX PD 25-NOV-1987.
XX PF 01-JAN-1987; 87EP-0200690.
XX PR 01-JAN-1987; 87EP-0200690.
XX PR 24-JUN-1983; 83US-0507419.
XX PR 29-MAY-1984; 84US-0614491.
XX PR 29-MAY-1984; 84US-0614612.
XX PR 29-MAY-1984; 84US-0614615.
XX PR 29-MAY-1984; 84US-0614616.
XX PR 29-MAY-1984; 84US-0614617.
XX (GETH ) GENENTECH INC.
XX PA Bott RR, Ferrari E, Wells JA, Estell DA, Henner DJ;
XX PI WPI; 1987-328920/47.
XX DR N-PSDB; AAN71241.
XX
```

```
PT Bacillus strains not excreting subtilisin or neutral protease -
PT obtd. by recombinant DNA procedures, useful for enzyme prodn.
XX esp. of hydrolase(s) such as amylase
PS Example; Fig 7; 71pp; English.
XX B-subtilis 1168 chromosomal DNA was digested with EcoRI. A single
CC 6kb fragment hybridised to a fragment from the C-terminus of the
CC subtilisin structural gene in pS4 (see AAN71240). It was ligated to
CC EcoRI-cut pBS42 and the ligation mixture used to transform E.coli
CC ATCC 31446. Plasmid DNA was prepared from a pooled suspension of
CC transformant colonies and used to transform a protease deficient
CC strain of B-subtilis (BG84). Plasmid DNA from protease producing
CC colonies was digested with EcoRI and examined by Southern blot
CC analysis to isolate the 6kb fragment. A positive clone was identified
CC containing a plasmid designated pS168.1. Three HincII fragments
CC and a HincIII-EcoRI fragment were ligated into M13 vectors and
CC sequenced to obtain the entire subtilisin sequence from which the
CC amino acid sequence of subtilisin was deduced.
SQ Sequence 381 AA;
Query Match 6.0%; Score 337.5; DB 8; Length 381;
Best Local Similarity 26.8%; Pred. No. 1.4e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
QY 48 SSVVEYEVYVAFNGYFTAKARNSFISSALKSSSEVDNWRIPRNPSSDPSDFEVIQIK 107
Db 33 SSF--EKYIVGF-----KQTHSAMSSAKKQVI----- 59
QY 108 EKQKAGLLTLE-DHPNLIKRVTPQKVRSLKYAESDTPVCNETRWSQKWSRPLRRAS 166
Db 60 -SEKGGKVKQKQFYVNAATAATLDEKAVKELK---KDPFVA----- 95
QY 167 LSLGSGFWHATGRHSSRRLLRAIPROVAQTLQADVLWQMGYTGANVRVAVFDGLSEKHP 226
Db 96 -----YVEEDHIAHEYAQSPYGISQ-IPALHSGQYTGTSNKKVAVIDSGIDSSHP 146
QY 227 HFKNVKTNTWNTERT--LDDGLGHGTFVAGVIASMR---COGFAPDAELHIFRVTNN 281
Db 147 DL-NVRGGSFVPESETNPYQDGGSHGTHVAGTIAALNNSICVLGVSFASLIYAVKVLDT 205
QY 282 QVSYTSWFLDAFNAILKIDVLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAICND 339
Db 206 GSGQYSWIINGIEWAINNNMDVINNSLGGTGTSTALKTVVDKA---VSSGIVVAAAAGNE 262
QY 340 GP--LYGTLPNADQMDVIGVGIDFEDNTARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSGGSTSTVGYPAPKYPSTIAVAVNSNQRAFSSAGS---EL-----DYMAGVS 310
QY 398 VRGSGVKGCRALSGTSVAGPVVAVAGVTLVSTVQKRELVPASMKQALIASARRLPGVN 457
Db 311 IQSTLPCTGTCAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDLRESTATYIG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYVGKGLINVQAAAAQ 381

RESULT 21
AAP03737
ID AAP03737 standard; protein; 381 AA.
XX AC AAP03737;
XX DT 16-AUG-1990 (first entry)
XX DE Subtilisin gene product.
XX KW Carbonyl hydrolase; subtilisin; neutral protease; ds.
XX OS Bacillus subtilis.
XX
```

FH Key Location/Qualifiers
 FT Modified-site 1
 FT /label=fMET start codon.
 XX PN EP357157-A.
 XX PD 07-MAR-1990.
 XX PF 22-JUN-1984; 84EP-0202584.
 XX PR 24-JUN-1983; 83US-0507419.
 XX PR 29-MAY-1984; 84US-0614612.
 XX PR 29-MAY-1984; 84US-0614615.
 XX PR 29-MAY-1984; 84US-0614616.
 XX PR 29-MAY-1984; 84US-0614617.
 XX PR 29-MAY-1984; 84US-0614491.
 XX PA (GETH) GENENTECH INC.
 XX PI Bott RR, Estell DA, Ferrari E, Henner DJ, Wells JA;
 XX DR WPT: 1990-069509/10.
 XX DR N-PSDB; AAQ03536.
 XX PT Mutant prokaryotic carbonyl hydrolase enzymes -
 PT obtd. by site-directed oligo-nucleotide mutagenesis, used in
 PT food processing and cleaning industries.
 XX PS Claim 16; Fig 7; 39pp; English.
 XX CC Probe derived from subtilisin gene was used to isolate carbonyl
 CC hydrolase gene, mutant versions of which exhibit different oxidative
 CC stability and/or pH activity.
 XX SQ Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 11; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYEVIAVNGYFTAKARNFSISSALKSSSEVDNWRIRPNRPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKDDVI----- 59
 QY 108 EKQKAGLTLE-DHPNKKRVTPQKVFSLKYAESDPTVPCNETRWQKQSSRPLRRAS 166
 DB 60 -SEKGGKVKQKQFYVNAATAATLDEKAVKELK---KDPVA----- 95
 QY 167 LSLGSGFWHATGRHSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAFDTGLSEKHP 226
 DB 96 -----YVEEDHIAHEYAQSVPYGISQ-IPALHSGQGTGNSVAVIDSGIDSSHP 146
 QY 227 HFKNVKERTNWTNERT--LDDGLGHGTFTVAGVIASMR-----COGFAPDAELHIFRVFTNN 281
 DB 147 DL-NVRGGSFVSETPNPYQDSSSHGTHVAGTIAALNNSIGVIGVSPSASLYAVKYLDT 205
 QY 282 QVSTYSFLDAFNALKKIDVLNLSIGGPDFMD--HPFVDKWEVLTANNVIMVSAIGND 339
 DB 206 GSGQYSLINGIEWALSNNMDVINNLSLGGPTGTALKTVYDKA---VSSGIYVVAAGNE 262
 QY 340 GP--LYGTLANPADQMDVIGCGIDFEDNIARESSRGMTTWELPGCYGRMKPDIVTYGAG 397
 DB 263 GSSGSTGVGYPAKYPSTTAVGAVNNSNQASPSAGS---EL-----DVMAPGVS 310
 QY 398 VRGSGVGGCRALSGTSPVAVAGVTLVSTVQKRELVPNSMKOALITASARLPGVN 457
 DB 311 IQSTLPGGTGYGNGTSMATPHVAGAAALILS---KHPTWTNAQVRDLSTATYILG--N 365
 QY 458 MPEQGHCKLLOLLRAYO 473
 DB 366 SFYTGKGLINVQAAQ 381

RESULT 22
 AAR24131
 ID AAR24131 standard; Protein; 381 AA.
 XX AC AAR24131;
 XX DT 17-NOV-1992 (first entry)
 XX DE Bacterial serine protease mutant.
 XX KW Detergent; liquid; lipase; degradation; stable.
 XX OS Humicola lanuginosa.
 XX FH Key Location/Qualifiers
 FT Cleavage-site 106..107
 FT /note="cleaves signal peptide"
 FT Protein 107..381
 FT /note="mature protease"
 XX EP486073-A.
 XX PD 20-MAY-1992.
 XX PF 25-JAN-1991; 91EP-0200149.
 XX PR 14-NOV-1990; 90EP-0870212.
 XX PR 25-JAN-1991; 91EP-0200149.
 XX PA (PROC) PROCTER & GAMBLE CO.
 XX PI Johnston JP, Lenoir PMA, Thoen CAJK, Thoen CAJ;
 XX DR WPI; 1992-168685/21.
 XX PT Liquid detergent compsn. contg. lipase and protease - i.e.
 PT bacterial serine protease in which methionine near active site is
 PT replaced
 XX PS Disclosure; Page 9; 10pp; English.
 XX CC The sequence is that of a bacterial serine protease in which the
 CC methionine adjacent to the serine of the active site is replaced by
 CC another amino acid. The protease is used in a detergent
 CC composition which comprises conventional detergent ingredients plus
 CC an enzyme system comprising a lipase and the protease variant. The
 CC protease variant does not attack the lipase and thus prevents lipase
 CC degradation. Liquid detergent compsns. are obtd. which are stable
 CC during storage.
 XX SQ Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 13; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYEVIAVNGYFTAKARNFSISSALKSSSEVDNWRIRPNRPSSDYPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKDDVI----- 59
 QY 108 EKQKAGLTLE-DHPNKKRVTPQKVFSLKYAESDPTVPCNETRWQKQSSRPLRRAS 166
 DB 60 -SEKGGKVKQKQFYVNAATAATLDEKAVKELK---KDPVA----- 95
 QY 167 LSLGSGFWHATGRHSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAFDTGLSEKHP 226
 DB 96 -----YVEEDHIAHEYAQSVPYGISQ-IPALHSGQGTGNSVAVIDSGIDSSHP 146
 QY 227 HFKNVKERTNWTNERT--LDDGLGHGTFTVAGVIASMR-----COGFAPDAELHIFRVFTNN 281
 DB 147 DL-NVRGGSFVSETPNPYQDSSSHGTHVAGTIAALNNSIGVIGVSPSASLYAVKYLDT 205
 QY 282 QVSTYSFLDAFNALKKIDVLNLSIGGPDFMD--HPFVDKWEVLTANNVIMVSAIGND 339

Db 206 GSGQYSWIINGIEWAINNNMDYINMSLGPTGTALKTVYDKA---VSSGIYVAAAAGNE 262
QY 340 GP--LVGTNNPADQMDVLGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSGGSTVGYPAKYPSTTAVGAVNNSNQRAFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGYKGCRLSGTSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPVGN 457
Db 311 IQSTLPGGYGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYYGKGLINVQAAAQ 381

RESULT 23

AAR34463

ID AAR34463 standard; Protein; 381 AA.

XX AC

XX AAR34463;

XX DT 19-AUG-1993 (first entry)

XX DE Bacillus subtilis alkaline elastase.

XX KW Alkaline protease; detergent; stability; Bacillus; bleach; stains;
KW alkylbenzene sulphonates; cheese manufacture; dehairing enzyme;
KW leather preparation.

XX OS Bacillus subtilis.

XX PN W09307276-A.

XX PD 15-APR-1993.

XX PF 07-OCT-1992; 92W0-US08341.

XX PR 08-OCT-1991; 91US-0772087.

XX PA (CHEM-) CHEMGEN CORP.

XX PA (VIST-) VISTA CHEMICAL CO.

XX PI Fodge DW, Hsiao H, Lalonde JJ;

XX DR WPI; 1993-134465/16.

XX PT Alkaline protease produced by bacillus stable in alkaline
XX conditions - used in detergents and bleaches to decompose
XX proteinaceous stains

XX PS Disclosure; Page 34; 58pp; English.

XX CC Bacteria producing an alkaline protease that is stable at a pH
XX greater than 10 in the presence of linear alkylbenzene sulphonate
XX were isolated by collecting soil from an alkaline environment which
XX had been exposed to detergent contamination or which had a pH greater
XX than 12. Strains were isolated which had alkaline protease activity
XX which was stable in the presence of detergent. The protein
XX sequence was compared to that of Bacillus species alkaline elastases,
XX from B. carlsberg, B. subtilis and B. amyloliquefaciens BPN to
XX determine if the novel protease was from a bacillus species. The
XX novel protease showed between 60-75 percent homology to the elastase
XX from the other strains, thus indicating that the protease is from a
XX novel Bacillus strain designated Bacillus 164A. The alkaline protease
XX is used in heavy duty detergents as the enzymes are useful to decompose
XX proteinaceous stains and perform at high pH which are beneficial to the
XX detergency of the surfactants in the liquids. The enzyme is stable for
XX long periods in the detergent liquids and when stored at high pH. The
XX enzyme is also useful in cheese mfr. and as a dehairing enzyme for
XX leather preps. See also AAR34258-61 and AAR34463-6.

XX Sequence 381 AA;

XX SQ

Query Match 6.0%; Score 337.5; DB 14; Length 381;
Best Local Similarity 26.8%; Pred. No. 1.4e-16;
Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;

QY 48 SSTVVEYEVYAFNGYFTAKARNSFISSALKSSEVDNWRRIIPRNPSSDYPSPDFEVIQIK 107
Db 33 SST--BKKYIVGF-----KQTSAMSSAKKQVI----- 59
QY 108 EKQKAGLLTLE-DHPNLIKRVTPQKVFERSLKYAESDPTVPCNETRWSQKWQSSRPLRRAS 166
Db 60 -SEKGGKVKQKFYVNAATAATLDEKAVKELK---KDPVSA----- 95
QY 167 LSLGSGFWHATGRHSSRRLLRAIPROVAQTLOADVLWQMGYTGANVRVAYFDTGLSEKHP 226
Db 96 -----YVEEDHIAHEYAQSVPYGISQ-IPALHSGQGYTGSNVKVAVIDSGIDSSHP 146
QY 227 HFNKVKERTNWTNERT--LDDGLGCHGTFFVAGVTASMR-----COGFAPDABLHIFRVPTNN 281
Db 147 DL-NVRGGASFVFPSETNPYQDSSGSHGTHVAGTTAALNNSIGVLGVSPSASLYAVKVLDTST 205
QY 282 QVSYTSWFELDAFNAYAILKIDVLNLSIGGPDFMD--HPFVDKYWELTANNVIMVSAIGND 339
Db 206 GSGQYSWIINGIEWAINNNMDYINMSLGPTGTALKTVYDKA---VSSGIYVAAAAGNE 262
QY 340 GP--LVGTNNPADQMDVLGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSGGSTVGYPAKYPSTTAVGAVNNSNQRAFSSAGS---EL-----DVMAPGVS 310
QY 398 VRGSGYKGCRLSGTSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPVGN 457
Db 311 IQSTLPGGYGAYNGTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKLDLLRAYQ 473
Db 366 SFYYGKGLINVQAAAQ 381

RESULT 24

AAR74224

ID AAR74224 standard; Protein; 381 AA.

XX AC

XX AAR74224;

XX DT 10-JAN-1996 (first entry)

XX DE B. subtilis subtilisin.

XX KW Bacillus amyloliquefaciens subtilisin; carbonyl hydrolase; protease;
KW fusion protein; preproprotein; transport; cell membrane; B.subtilis;
KW autolysis; maturation.

XX OS Bacillus subtilis.

XX FH Key Location/Qualifiers
XX FT Peptide 1..106
XX FT Peptide /label= signal peptide
XX FT Peptide 107..381
XX FT Peptide /label= mature peptide

XX U55411873-A.

XX PD 02-MAY-1995.

XX PF 29-MAY-1984; 84US-0614612.

XX PR 01-APR-1986; 86US-0846627.

XX PR 29-MAY-1984; 84US-0614612.

XX PR 27-FEB-1990; 90US-0488433.

XX PR 11-AUG-1992; 92US-0928697.

XX PA (GEMV) GENENCOR INC.

XX XX

PI Adams RM, Power SD, Powers DB, Wells JA, Yansura DG;
 XX WPI; 1995-178127/23.
 DR N-PSDB; AAQ90042.
 XX Recovery of recombinant subtilisin mutants from host cells - by
 PT treatment with active subtilisin to cleave mutant from its
 PT pro-sequence.
 XX Disclosure; Fig 2; 32pp; English.
 PS The amino acid sequence of the Bacillus subtilis subtilisin protein.
 CC The gene is used in a method to produce a carbonyl hydrolase
 CC (subtilisin) e.g. the B.amyloliquefaciens subtilisin (AAQ90041) or other
 CC heterologous protein (produced as a fusion protein) e.g. human growth
 CC hormone, such that the desired protein is translated as a preproprotein
 CC which can be transported across the cell membrane but is not released as
 CC an enzymatically functional protein until the application of an external
 CC protease or a protease encoded by the host cell e.g. B.subtilis
 CC subtilisin or neutral protease (AAQ90043). The preproprotein sequence
 CC is mutated so that it is incapable of autoproteolytic maturation. The
 CC B.amyloliquefaciens sequence was mutated using the primers AAQ90044-5
 CC and 0*****, specifically at S221N, D32N, A48R or contained a deletion
 CC of 166 amino acids from the C-terminus of the protein.
 XX Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 16; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYIYVAFNGYFTAKARNSFISALKSSSEVDNWRIRPNRNPSSDYPSPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKKDVI----- 59
 QY 108 EQKAGLITLE-DHPNKRKVTQPKVRSRLKVAESDPTVPCNETRWQSQWQSRPLRRAS 166
 DB 60 -SEGGKVKQKQFYVNAATAATLDEKAVKELK---KDPSSVA----- 95
 QY 167 LSLGSGFWHATGHRSSRLRLRAIPQVAQTLOADVLWQMGYTGANRVAVDFDGLSEKHP 226
 DB 96 -----VVEEDHIAHEVAQSVFPGISQ-IPKALHSGQYTGSNVAVIDSGIDSSHP 146
 QY 227 HFKNVKERTWNTERT--LDDGLGHGTFFVAGVTASMBE---COGFAPDAELHIFRVFTNN 281
 DB 147 DL-NVKGASVFYSENPYQDQSSGHVAGTIALNNSIGVLGVSPASLAVKVLDTST 205
 QY 282 QVSYTSWFLDAFNAILKIDVLNLSIGGDFMD--HPFDKMWELTANNVIMVSAIGND 339
 DB 206 GSGQSWIINGIEWALSNNMDVINNSLGGTGTALKTVVDKA---VSSGIIVVAAAAGN 262
 QY 340 GP--LYGLTANPADQMDVIGVGGIDFEDNTAFSSRGMTWELPGGYGRMKPDIVTYGAG 397
 DB 263 GSGGSTVGYPAKYPSTIAGVYNSNQRASFSSAGS---EL-----DVMAPGV 310
 QY 398 VRGSGVGGCRALSGTSVSPVAGAVTLVSVTVOKRELVPNSMKQALIASARLPGVN 457
 DB 311 LGSTLPGGTGYGAYGNTSMATPHVAGAAALILS---KHPTWTNAQVRDRLESTATVILG--N 365
 QY 458 MFEQGHGKLDLLRAYQ 473
 DB 366 SFYTGKGLINVAQAQ 381
 RESULT 25
 RAY39229
 ID AAY39229 standard; Protein; 381 AA.
 XX AC
 XX AAY39229;
 XX 23-NOV-1999 (first entry)
 XX Bacillus subtilis subtilisin.

XX Subtilisin; protease; secreted protein; degraded; heterologous protein;
 KW catalytic triad mutation; autoproteolytic maturation; prosequence;
 KW fusion protein.
 XX Bacillus subtilis.
 OS Key Location/Qualifiers
 XX Peptide 1..106
 FT /note="Putative signal peptide"
 FT 107..381
 FT /label="Subtilisin"
 FT US5939315-A.
 PN 17-AUG-1999.
 PD 01-MAY-1995; 95US-04322279.
 PF 01-APR-1986; 86US-0846627.
 PR 29-MAY-1984; 84US-0614612.
 PR 27-FEB-1990; 90US-0488433.
 PR 11-AUG-1992; 92US-0928697.
 PR 01-MAY-1995; 95US-04322279.
 XX (ADAM/) ADAMS R M.
 PA (POWE/) POWER S D.
 PA (POWE/) POWERS D B.
 PA (WELL/) WELLS J A.
 PI Adams RM, Power SD, Powers DB, Wells JA, Yansura DG;
 XX WPI; 1999-526118/44.
 DR N-PSDB; AA206712.
 XX DNA sequence comprising subtilisin prosequence fused to a
 PT heterologous sequence or a catalytically inactive subtilisin
 PT variant, useful for the production of heterologous proteins
 XX Disclosure; Fig 2; 30pp; English.
 PS This sequence is the subtilisin protein of Bacillus subtilis. Subtilisin
 XX is an alkaline protease which Bacillus secretes. The protein sequence is
 CC used in the formation of a fusion protein which has an N-terminal
 CC subtilisin prosequence linked to a C-terminal non-subtilisin polypeptide.
 CC The C-terminal sequence can alternatively be a subtilisin which has an
 CC amino acid substitution in one or more of the catalytic triad of this
 CC sequence. The mutations take the form of a substitution at position 221
 CC in the wild type mature protein, where Ser is replaced by Ala. A
 CC substitution at position 32 in the wild type mature protein may also be
 CC used, with the wild type Asp replaced by Asn. A substitution may
 CC alternatively be made at position 64. These mutant subtilisins can be
 CC used as the second part of the fusion protein. The mutated versions of
 CC subtilisin are incapable of autoproteolytic maturation. The DNA fusion
 CC sequences are useful for the production of heterologous proteins. As the
 CC subtilisin mutants are incapable of autoproteolytic maturation, they are
 CC therefore bound to the Bacillus cell membrane. The inability of the
 CC mutant protein to leave the cell membrane means that they are not able to
 CC degrade desired proteins when heterologous protein production is being
 CC carried out in Bacillus cells. Heterologous proteins are sequestered at
 CC cell membranes and can be easily isolated and released by the action of
 CC enzymatically active subtilisin.
 XX Sequence 381 AA;
 Query Match 6.0%; Score 337.5; DB 20; Length 381;
 Best Local Similarity 26.8%; Pred. No. 1.4e-16;
 Matches 117; Conservative 66; Mismatches 156; Indels 97; Gaps 17;
 QY 48 SSTVVEYIYVAFNGYFTAKARNSFISALKSSSEVDNWRIRPNRNPSSDYPSPDFEVIQIK 107
 DB 33 SST--EKYIVGF-----KOTMSAMSSAKKKDVI----- 59

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:27:53 ; Search time 100 seconds
(without alignments)
2167.616 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNIWLLVLLVLCGKKH.....PRVKRPQLMQVHPKPTPSV 1052

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTRMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2915	51.9	552	4 Q8TAN4	Q8tan4 homo sapien
2	2231	39.7	952	5 Q9VP10	Q9vp10 drosophila
3	335	6.3	1239	16 Q9FB24	Q9fb24 streptomyce
4	346.5	6.2	1722	5 Q9ND22	Q9nd22 leishmania
5	337.5	6.0	379	2 Q9FDF4	Q9fd4 bacillus li
6	336	6.0	1102	2 P95684	P95684 streptomyce
7	335.5	6.0	1245	16 Q9PL54	Q9pl54 streptomyce
8	327.5	5.8	627	16 Q9RUD0	Q9rud0 deinococcus
9	324.5	5.8	379	2 Q45300	Q45300 bacillus li
10	322.5	5.7	374	2 Q9F942	Q9f942 bacillus li
11	322	5.7	382	2 Q45522	Q45522 bacillus sp
12	321	5.7	379	2 Q45301	Q45301 bacillus li
13	320.5	5.7	275	2 Q93L66	Q93l66 bacillus su
14	318.5	5.7	374	2 Q9F943	Q9f943 bacillus li
15	318.5	5.7	376	2 Q44684	Q44684 bacillus am
16	317.5	5.7	379	2 Q53521	Q53521 bacillus li

17	316.5	5.6	374	2 Q9F941	Q9f941 bacillus li
18	314.5	5.6	312	2 Q93Q20	Q93q20 clostridium
19	313	5.6	310	2 Q9FDF3	Q9fd3 bacillus li
20	313	5.6	382	2 Q87655	Q87655 bacillus su
21	310.5	5.5	1253	16 Q9FC06	Q9fc06 streptomyce
22	308.5	5.5	275	2 Q9R7J4	Q9r7j4 bacillus su
23	308	5.5	310	2 Q9FDF2	Q9fd2 bacillus li
24	304.5	5.4	323	2 Q45621	Q45621 bacillus sp
25	304	5.4	310	2 Q9F7C2	Q9f7c2 bacillus li
26	302	5.4	654	17 Q8UOC9	Q8uoc9 pyrococcus
27	297	5.3	412	2 Q9AER6	Q9aer6 thermoanaer
28	297	5.3	412	16 Q8RC68	Q8rc68 thermoanaer
29	294	5.2	321	2 P74937	P74937 thermoactin
30	293.5	5.2	322	16 Q9KAV3	Q9kav3 bacillus ha
31	291.5	5.2	640	2 Q934J3	Q934j3 prevotella
32	287.5	5.1	383	2 Q9KWR4	Q9kwr4 bacillus pu
33	287	5.1	379	2 Q45467	Q45467 bacillus sp
34	286.5	5.1	891	1 Q93635	Q93635 thermococcu
35	285.5	5.1	374	2 Q45523	Q45523 bacillus sp
36	285.5	5.1	379	2 Q45299	Q45299 bacillus li
37	285	5.1	1398	1 Q9P9L1	Q9p9l1 pyrococcus
38	284.5	5.1	561	16 Q8RBJ2	Q8rbj2 thermoanaer
39	278.5	5.0	379	2 Q66153	Q66153 bacillus sp
40	278.5	5.0	692	2 Q9EXK0	Q9exk0 bacillus ps
41	276.5	4.9	384	2 Q56365	Q56365 thermoactin
42	270.5	4.8	328	16 Q92Y41	Q92y41 rhizobium m
43	263.5	4.7	682	2 Q48674	Q48674 lactococcus
44	259	4.6	757	16 Q9K6G6	Q9k6g6 bacillus ha
45	258.5	4.6	397	2 P97097	P97097 bacillus sp

ALIGNMENTS

RESULT 1

ID	Q8TAN4	PRELIMINARY;	PRT;	552 AA.
AC	Q8TAN4			
DT	01-JUN-2002	(TReMBLrel. 21, Created)		
DT	01-JUN-2002	(TReMBLrel. 21, Last sequence update)		
DT	01-JUN-2002	(TReMBLrel. 21, Last annotation update)		
DE	Similar to membrane-bound transcription factor protease, site 1			
OS	Homo sapiens (human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
..RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BRAIN;			
RA	Strausberg R.;			
RL	Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC026330; AHH26330.1;			
KW	Protease.			
SQ	SEQUENCE. 552 AA; 61331 MW; FE28225645FFD6A8 CRC64;			

Query Match	51.9%;	Score 2915;	DB 4;	Length 552;
Best Local Similarity	99.8%;	Pred. No. 9.6e-205;		
Matches 551;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MKLVNIWLLVLLVLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSSTVWEYIVAF	60	
Db	1	MKLVNIWLLVLLVLCGKKHGLDRLEKKSFEKAPCGCSHLTLKVFSSSTVWEYIVAF	60	
Qy	61	NGYFTAKARNSFTSSALKSEVDNWRLLPRNNSSDYPSPDFEVIQIKKQKAGLLTLEDH	120	
Db	61	NGYFTAKARNSFTSSALKSEVDNWRLLPRNNSSDYPSPDFEVIQIKKQKAGLLTLEDH	120	
Qy	121	PNIKRYTPQKVFPSLKLYAESDTPVPCNETRWQSKWSSRPLRRASLSLGSFGFWHATGRH	180	
Db	121	PNIKRYTPQKVFPSLKLYAESDTPVPCNETRWQSKWSSRPLRRASLSLGSFGFWHATGRH	180	
Qy	181	SSRRLRAIPROVAQTLQADVLWQMGYTGANVAVFDTGLSEKHPHFKNKERTWNTE	240	
Db	181	SSRRLRAIPROVAQTLQADVLWQMGYTGANVAVFDTGLSEKHPHFKNKERTWNTE	240	

```

Db 181 SSRLLRAIPROVAQLOADVLWQMGTYGANVRKAVFDTGLSEKHFKHVKVKNRTNWTNE 240
Qy 241 RTLDGGLGHGTGVAGVIASMRCEGAFAPDAELHIFRVTNNQVSYTSWFLDAPNFYAILKK 300
Db 241 RTLDGGLGHGTGVAGVIASMRCEGAFAPDAELHIFRVTNNQVSYTSWFLDAPNFYAILKK 300
Qy 301 IDVLNLSIGGPDMPDHPFVDKVKWELTANNVIMVSAIGNDGPLYGLTLNADQMDVIGVGG 360
Db 301 IDVLNLSIGGPDMPDHPFVDKVKWELTANNVIMVSAIGNDGPLYGLTLNADQMDVIGVGG 360
Qy 361 IDPENTARSSRGMTTWELPGGGRMKPDIVTYGAGVRGSGVGGCRALSGTGSVASPW 420
Db 361 IDPENTARSSRGMTTWELPGGGRMKPDIVTYGAGVRGSGVGGCRALSGTGSVASPW 420
Qy 421 AGAVTLVSVQRELVPASMKQALIASARLPQVNVMEFQGHGKLLDLRAVQILNSYKP 480
Db 421 AGAVTLVSVQRELVPASMKQALIASARLPQVNVMEFQGHGKLLDLRAVQILNSYKP 480
Qy 481 QASLSFSDIDTFCPTWPKYCSOPIYGGMPTVNVYTLNGMGTGRIVDKPDWQPLPQ 540
Db 481 QASLSFSDIDTFCPTWPKYCSOPIYGGMPTVNVYTLNGMGTGRIVDKPDWQPLPQ 540
Qy 541 NGDNIEVAFSFS 552
Db 541 NGDNIEVAFSFS 552

RESULT 2
Q9VP10
ID Q9VP10 PRELIMINARY; PRT: 952 AA.
AC Q9VP10;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG1169 protein.
GE CG1169.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Fandell M.D., Zhang Q., Chen L.F.,
RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Besson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler K., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Murphy B., Murphy L., Nusskern D.R., Pacle J.M.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

```

```

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Rao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smitch H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AE003595; AAF51752.2;
DR HSP; P00782; 150P.
DR FlyBase; FBgn0037105; CG1169.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 2.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 952 AA; 107683 MW; 0391F95C89BD1D42 CRC64;

Query Match 39.7%; Score 2231; DB 5; Length 952;
Best Local Similarity 44.5%; Pred. No. 2.8e-154;
Matches 445; Conservative 147; Mismatches 230; Indels 178; Gaps 17;

Qy 47 ESSTVVEYEVIVAFNGYFTAKARNSFISALKSESDVNNRIIPRNPSSDYPDFEVIQI 106
Db 19 FTAIVVNEPIVHFPHSKYFAPVRESYIAAKLGSNTNRIIPRLNLAQWYFSDILRV 78
Qy 107 KEKQAG-----LITLEDHFNKIKVTPQKVFSLKY-AESDPTVPCNETRWSQKQSRP 161
Db 79 CDGYESSEFIETLQTHFSVAVVQKSVRRILNDAYSNTL----- 121
Qy 162 LRRASLSGGFHWATGRSSRLRLRAP-----RQAQTLQADVLWQMGTYGANVAV 216
Db 122 -----YIH---RHQGVLRNPNNDNRHQLCSVLHANILMLKIGITGKGVKVI 167
Qy 217 FDTGLSEKHPHFNKIKVKNRTNWTNERTLDGGLGHGTGVAGVIASMRCEGAFAPDAELHIFR 276
Db 168 FDTGLTKHHPHFNKIKVKNRTNWTNERTNKSLLDRVSHGTFVAG----- 206
Qy 277 VFTNQVSYTSWFLDAPNFYAILKKIDVLNLSIGGPDMPDHPFVDKVKWELTANNVIMVSAI 336
Db 207 -----VSYTSWFLDAPNFYAILKINILNLSIGGPDMPDHPFVDKVKWELTANNVIMVSAI 260
Qy 337 GNDGPLYGLTLNADQMDVIGVGGIDFEDNIARESSRGMTTWELPGGGRMKPDIVTYGA 396
Db 261 GNDGPLYGLTLNADQMDVIGVGGIDFEDNIARESSRGMTTWELPGGGRMKPDIVTYGS 320
Qy 397 GYRGSYKGGCRALSGTGSVASPVAGVATLWS-TVQKRELVPASMKQALIASARLPQ 455
Db 321 QVGSQVDRKGCRRLLSGTGSVASPVAGVATLWS-TVQKRELVPASMKQALIASARLPQ 380
Qy 456 VNMFEQGHGKLLDLRAVQILNSYKQASISPSYIDLTECPYMPYCSOPIYGGMPTVNV 515
Db 381 YNMFQAGAGKLNLLKSNQILLSSYKPKITLIPALDFTQ-NYMPYSSQPLYYGSSVAIA 439
Qy 516 VTILMGVTCRIVDKPDWQPLPQNGDNIEVASYSVSLVMPWSGYLAISVTKKAASW 575
Db 440 VTILMGVTSVHIVGIPKIPDEFENQGFQVSAQVSPVYVFWPTWGMVSFVIAKKEGF 499
Qy 576 EGIAGQHVMTVASPATESKNGAETSTVKLPKIKVLIITPPRSKRVLDYQNLKRYPP 635
Db 500 EGVCQSGTIVLVESFKQNET---HVTEDFPITIKVTKPRNKRILMDWTHSLRPP 556
Qy 636 GFYPRDLNRKMDPLDMNGDHTHTNFRDMYQHLRSMGYFVEVLGAPCTFCDASQYGLIM 695
Db 557 RYTPRDDLLKVLDPDLWRADHITNFRDMYTHLRNVGYIDVLEPFTCFNASDYGALLI 616
Qy 696 VDSSEYEPPEIAKLREDV-DNGLSLVIFSDMTYSYVRKVKFYDENTQWMPDPTGGAN 754
Db 617 VDPERGFDEEINALQENYKRGVNVVFGDWYNTVTKKIKFEDENTQWMPDPTGGAN 676
Qy 755 IPALNELLSVNMVGFSDGLYEGEFTLANHDMYASGCSIAKFPED--GVVITQTFKDOGL 812

```

Db 677 IPALNDLLKPGFAGDFVGEHGFGLGDHSMYASGATIVKFFPMNPGDIIIVGTKLNDQGL 736
QY 813 EVLKQET---AVVENVPILGLYOIPA----- 835
Db 737 SIINSKTPSKVAKLDVPFQFQKANSIQSNEEIVVAESNLAEALPTDYSTFKNRVLL 796
QY 836 -----EGGRIVLYGDSNCLDSDSHRQKDCFWLLDALLQYTSYGVTTP 877
Db 797 LRTKQRSIPAKSNHETKNEGRIVAVYGDNSCLDSTHLEKACYWLLITFLDFALN----- 851
QY 878 SLSHSGNRQPPSGAGSVTERMEGNHLHRYSKVLEAHLGDPKPRPLPA----- 926
Db 852 --SH-----KSLQLNLNRTFHKLERAPLPLRISQSIKRSQD 890
QY 927 --CPRLSWAKPQPLNETAPSNLWKHOKLLSIDLKKVLPN 964
Db 891 NNCEQFKWLAPTQKNA-----ERKSSIIVILEN 922

RESULT 3
Q9FBZ4 PRELIMINARY; PRT; 1239 AA.
AC Q9FBZ4
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative secreted peptidase.
GN SCO7188 OR SC8A11.16C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RN Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lark L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RL coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL391041; CAC01588.1; -.
DR HSSP; Q99405; 1MPT.
DR InterPro; IPR003137; PA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1239 AA; 128505 MW; 8F5E9AC68EB1260A CRC64;
Query Match 6.3%; Score 355; DB 16; Length 1239;
Best local similarity 22.8%; Pred. No. 9.3e-17;
Matches 217; Conservative 116; Mismatches 363; Indels 254; Gaps 45;
QY 203 WQMGYTCANRVAVDFDGLSEKHPFKN-VKERTNNTNERTLDGLGHGTFVAGVI----- 257
Db 236 WAGGNTGGGEVAVLDITGVDAHPDLADRIAAROSFVPDENTDDRGHGTHVASTIAGTG 295
QY 258 -ASMRECGQFAPDAELHIFRVFTNNQVSYSWFLDAFNALIKK-IDVLNLSIGGPFMD 315
Db 296 AASAGKEGVAPGARLSIGKVLDSGRQISWTLAAMEWAVERHAKIVNLSLGSQSD 355
QY 316 --HPFVDKVMELTA-NNVIMVSAIGNDGPYGLTNNPADQMDVITGVGGIDFEDNIAPSS 372
Db 356 GSDPMSRAVDRLSAQTGALFVVAAGNGEA-GSIGAPGVATSAITVGAVDATDTLAPSS 414
QY 373 RGMITWELPGYGRMKPDIVTYGAGVR-----GSGVKGGRALSGLSVSPVAVAGVTL 426
Db 415 QG-----PRVDGALKPEITAPGVGILAAANSFSAAGNGAYQSLSGTSMATPHVAGAAAL 468
QY 427 LVSTVQKRELNVNPMKQALIASARRLPGVNMFQGHGKLDL---LRA--YQILNSYKPK 481
Db 469 LAAA---RPDLSGSALKDVLASSHRTPRYDAFOAGSGRVDVDAVRAGVYASATAPG 525
QY 482 ASLSPSYIDLTECPYMPYCSQPIYYGGMPTVNVNITLNGMGVTRIGRVDKPDWQPYLPQN 541
Db 526 SSGP-----VRLVYTYTNTGAATLLELSVAATHAP----- 557
QY 542 GDNLEAVF--SYSSVLMPWSGYLAISIVTKKAASWEGIAAGHVMITVASPAETESKGA 599
Db 558 ----EGVFRLSASRVTPAHGTADVTITDGGSGAGRAYSGQILAT-----DADARVA 608
QY 600 EQTSTVKLPIKVKII-----PTPPSRKRVLDQYHNLRYPPG-----YFPRDN- 642
Db 609 -HTAVSAGPVVRHKLTVHFKDADGNPVGVDLLKSGDSESLPVLVGDSGTAELYPEDTY 667
QY 643 --LRMKNDDPLDWDGDIHTNFRDMYQHLRSMGYEVEVLGAP-----FTCFPASQYGT 692
Db 668 SALAFKTVP-GVHGPH-----SWGMALLGDPEVRLTEDTAVTFDASR--- 708
QY 693 LLMWDSSEYFPE--ETAKLRDDVDNGLSLVIFSDWYNTVMRKVKFVDENTRWMPDT 750
Db 709 ---VERIETTVPQTEATYQRLDYQSRMG-----GTYTGTLETQTA-YDS---LWAOPTT 757
QY 751 GGA-----NIPALNELLNVNMNMGFSGLYEGETLANHDMYASGGSIAKFP 797
Db 758 HKYTHGDFLYNARWRKEQPALT--VSTRTDFTDVLROG-----GVTPALP 800
QY 798 EDGVVITQTFKDOGLE-----VLKQETAVVENVPILGLYOIPAGGGGRIVLY 844
Db 801 KGTRTLPLVFAGDAAAEYARLDARGAVVVRDDVDAD-----GVQAAANVAAGATLLL 855
QY 845 GDSNCLDSDSHRQKDCFWLLDALLQYTSYGVTTPSLSHSGNRQPPSGAGSVTPERMGNH 904
Db 856 VVNN--EDGR-----ALRCY-----GEPFGPPVALDVALLSTDEGEK 890
QY 905 LHRYSKVLEAHLGDPKPRPLPACPRL-----SWAKPQPLNETAPSNLWKHOKLLSIDL- 957
Db 891 LAAQAKVARGV---TVTSRPVSPYVYDLLASWNEIPTMTSRAD---SRSLARVDVAF 944
QY 958 -----DKVVLNFRSNRPQVRPLSPGESGAMDIPGIMGRVNOEVGOTIP 1003
Db 945 DSQLPGSGGGEFRYDWPVPGSGWTFGGQPEPVS-GTRTDWYSTGDY---RWNNQ----- 994
QY 1004 VFAFLGAMVVLAFVVOIN-KAKSRPKRRPR-----VKRPQLMQOVHPP 1047
Db 995 --AYAGGV-----IYEIGAKTAYRPGSRQSEWEFGPVRPHLANDAYRSP 1036

RESULT 4

Q9NDZ2 PRELIMINARY; PRT; 1722 AA.

AC Q9NDZ2; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Possible serine protease.

OS Leishmania major.

OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

NCBI_TaxID=5664;

SEQUENCE FROM N.A.

STRAIN=FRIEDLIN;

RA Oliver K., Harris D., Ivens A.C., Quail M., Rajandream M.A.,

RA Submitted (May-2000) to the EMBL/GenBank/DBJ databases.

RL

[2]

SEQUENCE FROM N.A.

STRAIN=FRIEDLIN;

RC MEDLINE=98146435; PubMed=9477341;

RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,

RA Smith D.F.;

RT "A physical map of the Leishmania major Friedlin genome.";

RL Genome Res. 8:135-145(1998).

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 2.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Protease.

SQ SEQUENCE 1722 AA; 182741 MW; 8148DD78AC52DDIF CRC64;

Query Match 6.2%; Score 346.5; DB 5; Length 1722;

Best Local Similarity 17.4%; Pred. No. 6.7e-16;

Matches 250; Conservative 147; Mismatches 331; Indels 707; Gaps 53;

QY 166 SLSLGSGFW-----HATGRHSRRLL-----RAIPROVAQTQADVLQOM 205

DB 26 SLPLAEGPWLVSQAARRSTDAHAERPSIHPKRYLKYGTGTCIRNLTLGLVWSEDAERL 85

QY 206 G-YTCANVRVAFDTGLSEKHPKVNKERTNTWNTL-----DDGLGHGTFVA 254

DB 86 ACFSGRGVRVAVLDTGLC-----AGITESSR--NSVTCTSVVPGVACEDAGCAHGTSSV 137

QY 255 GVIA-----SMREQ-----GPAPDAELHIFRV-----TNNO 282

DB 138 SVLAQLSVSTPLPRESIAKETAWNRETHALADQFGLADPSIVRVLRIFDRQGTNR 197

QY 283 -----VSTSWFLDAFNAILK-----IDVLNLSIGGPDPMDFPV 319

DB 198 HLAALDLLREADEAGEAGRLHSTWNTSS---AHIRRDTEVDVLSISGSKDYHSSPV 254

QY 320 -DKVWELT-ANNVIMVSAIGNDGLYGLTLNADQMDVIGVGGIDFE----- 364

DB 255 QDRYRLMHEHGIVTVAAAGNDGVFGSVSPADMPGLAIGALRVGEHSRAQTQASR 314

QY 365 -----DNIAFSSRGMTTWELPGGYGRMKPDIVTYG-----AGVRGSGV 403

DB 315 TLGNLAGSLAAGGCHKSVAFHSGRPTTWELPFGACRAKPDILVALGQHWVAGVSAAAF 374

QY 404 KGGC-----RALSGTSVSPVWAGVTL----- 427

DB 375 SAAPRARSASSALQLRASGTSINAIPVAGVVALCLEAMSSTSVASAQTGGGVL 434

QY 428 -----VSTVOKRELVPASKQA----- 445

DB 435 DVQRNLSRVSHSLRVREALIRLTAIVPLEEAAASLPTWPRAPTPLHSDSSSTASNAHR 494

QY 446 -----LIASARRLPGVNMFEGQHGKLDLLRAYQLNSYKPAQSLSPSYDILTE-- 493

Db 495 RVLACVPLELYAGYLRLSRVLSLQSGAGEVQPIRA---LHAIVAKAAS--SAIDAPEAL 550

QY 494 ----- 493

Db 551 RSCASFAIPTCVIRIGYRIPRCSNSGDMSTSQSQPPHDPDGKAHGDGSLRPGCMGRAA 610

QY 494 ---CPTMWPYCSQPIYYGGMPTVVNYI----- 518

Db 611 PPAAYWPFSDQAVYFCATPVLLINLSLHLCPPSSSAATDQADAAAAAAGEAARRGTT 670

QY 519 -----LNG-----MGVTG 526

Db 671 HDHVTYVITKVGSRRLAREGHPSRNSSHQNDACPLLLFEGAENRHRVGSTPTRELLNITA 730

QY 527 RIVDKP-----DWQPYL----- 538

Db 731 RSASPAACTDGEVVGGEATVTGPRHTQAGVCSFAKRRSTQQRWMOHLLRVATELTVVAS 790

QY 539 ---PONGDNIE-----VAFSYS-SVLWPWSGYLAISIVTKK----- 571

Db 791 RSPTRANVVEQLSPPTSFSLSVAVSSPASATHLCYDVARETVAVWRKKGRKSPSVYK 850

QY 572 -AASW-----EGIAOGHYMITVASPAETES-----KNGAEQIST 604

Db 851 AASRSADNNDGHDQHTDGAQVPADRCVPLFLPRMLSVGALHIFTGGSSQPT 910

QY 605 VKLPIKVIIPTPPRSKRVLWD-----QTHNLRYPPGYFPDNLRMKNPLD--- 651

Db 911 LTVPEAVCVWEPPRVORVLIDTSLDNWNPNTATSNL-FTAGDDPHES---DTDGVDAAL 966

QY 652 -----WNGDHTHTNPRDYOHLR-SMGYFVEV----- 677

Db 967 RRGQRQRHERAYAEASGGDHPYTNLALLYLRLHTLGMVETFFLLHMTTITTSIAANG 1026

QY 678 ---LGAP-----FTCFDASQ-----YGTLLMVDSEEEYFPEETAKLRDV----- 714

Db 1027 SSPLPAPPAEHTFNASHAAQAPAGALAHVGTLLIIVDPERPLTHGMRRLITRAVLGGADR 1086

QY 715 -DNGLSLVIFSDYNTSVMRKV-----KFYDENTROWMPDT----- 750

Db 1087 SADGLDVLVTDYTSADLAQLHWARDKSLDNAERSRISGTDAADEAVGVDGRDAVRE 1146

QY 751 -----GGANIPALNELLS-----VNMNFGDGLY 774

Db 1147 LRALRQLENGSTRGLAGSHVPSWNRWLSEVTVASGSAAANSNDCAAPLSEGLLDGKS 1206

QY 775 EGFEFTLANHDMYASGCSIAKFPEDGVVITQTF-----KQGLVLEKQETAVVENVPTIG 829

Db 1207 ELPPFELSENIVI-----DCVVVDAAAPTGRSEPGGNASASHGGTITSKAVYA 1254

QY 830 L-----YQIPAEG-----GGRIVLYGDSNCLDSDSHRQKDCFLLDALLOYTS 871

Db 1255 LRSGLQNLNAAQVLRWRLPQAQAAQRAAQAAGVASKGVHESN-----VHSSG 1302

QY 872 YGVTPPSLSHSGNRQRPSPGAGSVTPPERMEGNHLHRYSKVLEAHLGDKPRLPACPLRS 931

Db 1303 HGTC-RIGHASNT-----SAVPSQEG-----EAVICNVP-----G 1333

QY 932 WAKPQPLNETAPSNLMKHKLLSIDLKVLPNFRSNRPQVRPLSPGSGAWDIP 986

Db 1334 WYQQQ-----RRLVKRTIVSTRDSSAPAAVYEDGWEDVP 1367

RESULT 5

Q9FDF4 PRELIMINARY; PRT; 379 AA.

AC Q9FDF4; 01-MAR-2001 (Tremblrel. 15, Created)

DT 01-MAR-2001 (Tremblrel. 15, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Kera.

OS Bacillus licheniformis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;


```
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-OWU 1411T;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RL Bacillus licheniformis OWU 1411T."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282893; AG00492.1; -
DR HSSP: P00780; 1CSE.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE.ASP; 1.
DR PROSITE: PS00137; SUBTILASE.HIS; 1.
DR PROSITE: PS00138; SUBTILASE.SER; 1.
SQ SEQUENCE 379 AA; 38904 MW; DC65F6A93EBC69CC CRC64;

Query Match 6.0%; Score 337.5; DB 2; Length 379;
Best Local Similarity 25.7%; Pred. No. 2.5e-16;
Matches 126; Conservative 66; Mismatches 169; Indels 129; Gaps 17;

QY 1 MKLVNWLVLVLLVLCGKKHLGRLEKKSFEKAPCPGCSHLTLKVFSSV-----V 52
Db 2 MRKKSFWLGLTALM-----LVFTMAFSDSASAAQPAKNV 36

QY 53 EYEYIVAF-NGYFTAKARNSFISALKSSVD-NWRIIPRNPSSDYPSPDFEVIQIKEQ 110
Db 37 EKDIYVGFSGVKVTASVKKDIKES--GGKVQKQFRIINAAKALD----- 80

QY 111 KAGLLTLEHPNLIKRVTPORKVPRSLKIAESDPTVPCNETRWSQKWSSRPLRRASISLG 170
Db 81 KEALEEVKNDPV-----AYVEED----- 99

QY 171 SGFWHATGRHSSRLRLRAIPRQVAQLQADVLQMGYTGANRVAVFDGLSEKHPHEKN 230
Db 100 -----HVALAQAQTPYGT-PLIKADKVOAQYKGANRVAVLDGTGIQASHPLNV 149

QY 231 VKERTNWTNERTLDDGLGHGTFVAGVIASMQEQ---GPAPDAELHIFRVFTNQVSYTS 287
Db 150 VGGASFVAGEAYTWDGNGHTHVAGIVAAALDNTGVLGVAPNVSLVAVKVLNSGSGSYS 209

QY 288 WFLDAFNAILKIDVLNLSIGGDFMD--HPFVDKRWELTANNVIMVSAIGNDGPI--Y 343
Db 210 GIVSGIEWATTNGMDVINLSLGGSPSGSTAMQAVDNAY---ARGVVVAAAGNSGSGNT 266

QY 344 GTLNNADQMDVIGVGIGDFEDNIARFSSRGWTTWELPGYGRMKPDIVTYGAGVRGSGV 403
Db 267 NTIGYPAKVDVIAVAGVDSNRSRFSVGA---EL-----EVMAPGAGVYSTYP 314

QY 404 KGCRLALSGTSVSPVAVGAVTLVSTVOKRELNVNPSMKQALIASARRLPGVNMPEQGH 463
Db 315 TSYATLNGTSMASPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG---SSFYKG 369

QY 464 GKDLLLRAYQ 473
Db 370 GLINVEAAQ 379

RESULT 6
P95684 ID P95684 PRELIMINARY; PRT; 1102 AA.
AC P95684;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Subtilisin-like protease.
OS Streptomyces albogriseolus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-S-3253;
RX MEDLINE-97144528; PubMed-8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
RL Streptomyces albogriseolus."
RL J. Bacteriol. 179:430-438(1997).
DR EMBL: D83672; BAA12040.1; -
DR HSSP: P00782; 2SBT.
DR InterPro: IPR002860; GH_BNR.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF02012; BNR; 2.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE.ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE.HIS; 1.
DR PROSITE: PS00138; SUBTILASE.SER; 1.
KW Protease.
SQ SEQUENCE 1102 AA; 114128 MW; F9E4AD2590FE59E CRC64;

Query Match 6.0%; Score 336; DB 2; Length 1102;
Best Local Similarity 22.7%; Pred. No. 1.9e-15;
Matches 164; Conservative 107; Mismatches 260; Indels 190; Gaps 33;

QY 203 WQMGYTGANRVAVFDITGLSEKHPHF-KYKERTNWTNERTLDDGLGHGTFVAGVIA--- 258
Db 209 WEAGYDGKGVKIAVLDTGVDATHPDLKGQVTSKNTFSAPTTGVDVVGHTHVASIAAGTG 268

QY 259 --SMRECQGPAPDAELHIFRVFTNQVSYTSWFLDAFNAILKIDVLNLSIGGPDPMDH 316
Db 269 AOSKGYKGVAPGAKLNGKVLDDAGFGDDSGILAGMEWAAAGADIVNNSLGG---MDT 325

QY 317 PFVD---KVELTA-NNVIMVSAIGNDGLYGTLLNPAQDMQDVGIGGIDFEDNIARFS 371
Db 326 PETDPLEAAVDKLSAEKGLIFALTAAGNEP--OSIGSPGSADSALTGVAVDDKDLADFS 383

QY 372 SRGMTWELPGYGRMKPDIVTYGAGVRGSGVGK-----GCRALSGTSVASPV 419
Db 384 STGPRL-----GDGAVKPDLTAPGVDITAAASAKGNDIAKEVGEKPGACMYTISGTSMATPH 438

QY 420 VAGAVILLVSTVOKRELNVNPSMKQALIASARRLPGVNMPEQGHGKLDILRAYQILNSYK 479
Db 439 VAGAAALL---KQHPPEWYAEKLGALTASTKD--GKYPPEQSGRVRQVDKAITQTVAIE 494

QY 480 PQASLSFSYIDLTECPYMWPYC-----SOPILYGGMPTVNVNITLNGMGTGRIVDKPDW 534
Db 495 P-----VSLSEGVQVQHPHADDKPVTKKLTVRNLTG-EDVILKLTSTATG----- 537

QY 535 QPYLPQNGDNIEVAF---SYSSVLMPWSGYLAISISV-TKKAASWEGIAQGHVMTVASP 590
Db 538 -----PKGKAAPAGFFTLGASTLTVPANGTASVDVTDATRLGGAVDGTYSAYVVATGAGQ 592

QY 591 A-----ETSKNGAETSIVKLPKIKVILPTPPRS-----KRVLMD 626
Db 593 SVRTAAAVEREVESYN-----VTLKVLDRSGKATANYMAYLSGLTGLGKDRSYA 641

QY 627 QYH-----NLRYPPGYFPDR-NLRMKNDPLDWDGDIHTNFRDMYQHLRSMGYEVEYVGA 680
Db 642 PYEADGAVSVRVPKGGYVLDASVILGADPTWRG-----ADWLAQ 681

QY 681 PFTCFDAQSYGTLMLVDSEEE-----YFPEEIAKLK-----RDVNGLSLVIFSDWYNTS 730
Db 682 P--KLDVTR-NTTVVDARKAKPVKVTVPCKAQAQFASADYTITETNDSAVSYG----- 732

QY 731 VMKRVFYDENTROWMPDGTGGANIPAL-----NELLSVNMGFSGLGEGETLANHD 784
Db 733 -----WLENYSGFRSAHLGPGQITNGTLSQQWNTHFSNGA-KAQYTA----- 773

QY 785 MYVAGSGSIKFPEDGVVITQTFKDOGLEVLKOSTAVVENVPILGLQIPAEQGGRIVLY 844
Db 774 ---ISGKVKKL---ATGYTRAFK-----AKEFATVQ-----VGMGAASGKKGAVTAF 816

QY 845 G 845
```

Db 817 G 817

RESULT 7

Q9RL54 PRELIMINARY; PRT; 1245 AA.

AC Q9RL54; 01-MAY-2000 (Tremblrel. 13, Created)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Probable secreted peptidase.

GN SC00432 OR SCF51A.10

OS Streptomyces coelicolor.

OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Murphy L., Harris D.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2);

EX MEDLINE=97000351; PubMed=8843436;

RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,

RT Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RL the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

RN [4]

RN SEQUENCE FROM N.A.

RC STRAIN=A3(2); / M145;

RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RT coelicolor A3(2).";

RL Nature 417:141-147(2002).

RL EMBL; AL121596; CAB56662.1; -

DR HSP: P00782; 1SUE.

DR InterPro: IPR003137; PA.

DR Pfam: PF02225; PA; 1

DR Pfam: PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILASE_HIS.

DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

DR PROSITE; PS00136; SUBTILASE_SER; 1.

SQ SEQUENCE 1245 AA; 130895 MW; 74EE92DB9C4DDE60 CRC64;

Query Match 6.0%; Score 335.5; DB 16; Length 1245;

Best Local Similarity 21.5%; Pred. No. 2.5e-15;

Matches 218; Conservative 116; Mismatches 350; Indels 331; Gaps 42;

QY 159 SRLRSLSLGSGFWH-----ATGRHSSR-----RLLRAPROVAQTQADVLW 203

Db 169 ARAVVDDEALGR-FWKRLVPAAGQRAKAAATPRVWLDGRVGSASLDRSPAQIGAPDV-W 226

QY 204 QMGYTGANRVAVFDVTLGSEKHPHF-KVKERTNMWNTLDDGLGHGTGTVAGVI----- 257

Db 227 SAGYRGDGVKVAVLDTGADQSHPLAGRAAAKDFSGSGSGTNDVFHGCHTIVASIVGGSGA 286

QY 258 ASKRECQGFAPDAELHIFRVFTTNQVSYTSWFLDAFNAYLKKIDVLNLISGCPDFMD-- 315

Db 287 ASGSRQGVAPAKRLVGVKVGDDGSGESQVIAGMEWAADQAGADVNNISGSSGATDGT 346

QY 316 HFYDKWELT-ANNVIMVSAIGND---GPLYGLNNPDOMDVGICVGGIDEDFNIAFIS 371

Db 347 DPMQALNDLSRRGTGLFYVAAGNEGEGP--RTVSGPGAADAALTVGAVDRDLSAPFS 404

QY 372 SRGMTTWELPGTGRMKPDIVTYGAGVSGVKGCC-----RALSGTSVASPVVAG 422

Db 405 SRGPRL-----GDVAVKPDVTAGVGIVAAARAAGSAMGDPVDEHYHTAASGTSMATPHVAG 459

QY 423 AVTLIVSTVQKRELVPASMKQALIASARRLPGVNMFEQGHGKLDLLRA----- 471

Db 460 AALL--AORHPDWTGAOLKDALISTAVTDGQKVTQGGGRIDVRAAGLAVTATGTL 516

QY 472 -----YQILNSYKPAQL----- 484

Db 517 VMGPFTSRDTEPVTSRVRTNSDEDVTLSLAVELATEGCKAPAGSARLGSDSVRVPAG 576

QY 485 SPYSIDLTECPYMPYCQPIYGGMPVWVNTILNGM-----CVTCRI 528

Db 577 SSAEVLVDP--ARAGQKFG--YVTATTADGSAVAHTTSLVVHGHTRITVRT 629

QY 529 VDKPDWQYLPQNDNIEVAFSYSVLWPMWSGLAISVYTKKAASWEGIAQGHVMTVA 588

Db 630 IDK-----DGEQV-----ADLPTINGAEGFVGYSDPAAVAEVEGVIQNTSWI-- 674

QY 589 SFAETESKNGAEQTSIVKLPK-----KIPTPRSKRVKRWQYHNLRYPPGY 637

Db 675 ----TSAEDGELRHVLPKVTMDVTLDAKTVFVEITPR----- 715

QY 638 FPRDLNRMKNDPLDMNGDHIHNFND-----MYQHLRSMGFVEV----- 677

Db 716 -----PABQRIISYQTYREIGRSLTGTMYFDIAKRLYISPTAKVTEGTEF 764

QY 678 -----LGAPF-----TCFDASQY-----GTLIMVDS----- 698

Db 765 ASRWOLVAPLEAKVSRVTFGLGAYMPASLPFERGATLTVADGVAADAFSRRGRV 824

QY 699 ----EEEYEPPEIAKLRRDNGLSLVIFSDMYNTSVMRKVFYDENTROMNPDTGGAN- 754

Db 825 AVVDNPTGAGERALVERAAAGVSAVILVHFNDFGTR-----WRP-TGDRNA 871

QY 755 IPALN-----ELLSVWNNMGFSDGLYEGETLANHDMYYVASCSTAKFP-----E 798

Db 872 LPTLRIGKGTGAELLARIARGSTSVRFSG--TARSPYLVDVMQVSEQVPRKVVHVSAR 929

QY 799 DGWVITQPKDQGLEVLKQETAV----VENVPILGLYOIPAEGGGR--IVLYGDSNCLDD 852

Db 930 NSAVVRTTYADNGVTGWASEQRFGRWPYQNTAWLQYTRFVPLGFERTVVSAGDTAWOHR 989

QY 853 SHROKDCFWLLDALLOYTSYGVTPPSLS-----HSGNRQ-----RPSGAGS 894

Db 990 VHMT-----TFVDVQPLIAGMRDPRTYAGVRPRETWOGAVVRPSIPAGT 1036

QY 895 VTPERMEGNHLH-RYSKYLEAHLGDPKRPPLPACPRLSWAKPOPLNETAPSILWK 948

Db 1037 TPTVTRDGNVLRLVAETDSQAG-----HNSPPSG-GDTASAVLYLR 1077

RESULT 8

Q9RUDD PRELIMINARY; PRT; 627 AA.

ID Q9RUDD; 01-MAY-2000 (Tremblrel. 13, Created)

AC Q9RUDD; 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Serine protease, subtilase family.

GN DR1459.

OS Deinococcus radiodurans.

OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;


```

DE Subtilisin precursor (EC 3.4.21.62) (Fragment).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=NCIMB 10689;
RX MEDLINE=20561724; PubMed=11109488;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
RT feathers.";
RL Can. J. Microbiol. 46:1004-1011(2000).
DR EMBL; AF205190; AAG31027.1; -.
DR HSSP; P00780; ICSE.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 >374 SUBTILISIN.
FT NON-TER 374 374
SQ SEQUENCE 374 AA; 83390 MW; B61432A1B42C526E CRC64;

Query Match 5.7%; Score 322.5; DB 2; Length 374;
Best Local Similarity 25.2%; Pred. No. 3.1e-15;
Matches 122; Conservative 66; Mismatches 168; Indels 129; Gaps 17;

QY 1 MKLVNILLVLLCGKKHLDGRLKSEKAPCPGCSHLTKVFEFSIV-----V 52
   | : | | : | : |
Db 2 MRKSFVGLMGLTALM-----LVFTMAFSASASAQPAKRV 36

QY 53 EYEYIVAF-NGYFTAKARNSFISSALKSSEVD-NWRIIPRNPSDDYPSDFEVIQIKERQ 110
   | : | | | : | : | : | : | : | : | : | : | : | : |
Db 37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRIINAARAKLD----- 80

QY 111 KAGLLTLEDHPTNKRVTPQKRVFSLKYSDDTPVPCNETRWSQKSSRLRRASLSLG 170
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 81 KEALKEVKNDDPV-----ATVEED----- 99

QY 171 SCFWHATGRHSSRLRLRAIPROVAOTLQADVLQMGYTCANVRVAFTDTGLSEKHPFKN 230
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 100 -----HVAHALAQTPYGI-PLIKADKVOAQFGKCANVAVLDTGIQASHDLNV 149

QY 231 VKERTNWTNERTLDDGLGHGTFVAGVTASMKRECO---GFAPDAELHIFRFTNNQVSYTS 287
   | : | | | | | | | : | : | : | : | : | : | : |
Db 150 VGGASVFAGEAYNTDNGHGTHTVAGTVAALDNTTGLVGVAPSVSLYAKVLNLSGSGSYS 209

QY 288 WFLDAFNAILKIDVLNLSIGGDPFDND--HPFDVKWELTANNVIMVSAIGNDGPL--Y 343
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 210 GIVSGIEWATTNGMDVNMISLGGSGSTAMKQAVDNAY---ARGVVVAAAGNSGSGWT 266

QY 344 GILNPNADQMDVIGVGIDFDENTARFSSRGMTTWELPGYGRMKRPDIYTYAGVRGCV 403
   | : | | | | | | | : | : | : | : | : | : | : |
Db 267 NTIGYPKDYDVTAVGAVDSNRSASFSGA---EL-----EVMAPGAGVYSTYP 314

QY 404 KGCRLSGTSVSPVAVGAVTLVSTVQKRELVPASMKQALIASARRLPVGNMFEQCH 463
   | : | | | | | | | : | : | : | : | : | : | : |
Db 315 TTYATLNGTSMASPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSPYTK 369

QY 464 GKLDL 468
   | : | : |
Db 370 GLINV 374

RESULT 11
Q45522
ID Q45522 PRELIMINARY; PRT; 382 AA.
AC Q45522;

```

```

DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Prepro-subtilisin Sendai precursor.
GN APRS.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=G-825-6;
RX MEDLINE=95329264; PubMed=7605625;
RA Yamagata Y., Isshiki K., Ichishima E.;
RT "Subtilisin Sendai from alkalophilic Bacillus sp.: Molecular and
RT enzymatic properties of the enzyme and molecular cloning and
RT characterization of the gene. aprs.";
RL Enzyme Microb. Technol. 17:653-663(1995).
DR EMBL; D29698; BA006157.1; -.
DR HSSP; P29600; IGGI.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Signal.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 114 382 SUBTILISIN SENDAI.
FT NON-TER 382 382
SQ SEQUENCE 382 AA; 39901 MW; 9175A41FD5B9DE43 CRC64;

Query Match 5.7%; Score 322; DB 2; Length 382;
Best Local Similarity 27.4%; Pred. No. 3.5e-15;
Matches 124; Conservative 59; Mismatches 172; Indels 98; Gaps 19;

QY 27 KKSFKAPCPGCSHLTKVFEFSIVVEYE-----YIVAFNGYFTAKARNSFISSALK-SS 80
   | : | | | : | : | : | : | : | : | : | : | : | : |
Db 2 KKLTKVVASAA--LLLSISLTATSVSAEKKOYLIGFENQLQV---TEFVSSDKGOS 56

QY 81 EVDNRIIPRNPSDDYPSDFEVIQIKERKAGLLTLEDHPTNKRVTPQKRVFSLKVAE 140
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 57 ENSLFAEVNDESIEWELLYEFEDIPVSVVE---LSPEDVKDLEK-DP-----SITVIE 105

QY 141 SD--PTVPCNETRWSQKSSRLRRASLSLGSFGFWHATGRHSSRLRLRAIPROVAOTLQ 198
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 106 EDIEVTIINQVTPW-----GTR-----VQ 125

QY 199 ADVLQMGYTCANVRVAFTDTGLSEKHPFKNKTERTNWT-NERTLDDGLGHGTFVAGVI 257
   | : | | | | | | | : | : | : | : | : | : | : |
Db 126 ATATNRGITGIVRVAVLDTGIS-THPDL-NIRGYSVFPGPSYQDQNGHGTHTVAGTI 183

QY 258 ASMRE---CQGFADAEHLHIFRFTNNQVSYTSFIDAFNYAILKIDVLNLSIGGPDFM 314
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 184 AALNNSIGVGVAPNAELYAVKVLGANGSGSVSSIAQGLQWTAONNIHVNLSLGS----- 239

QY 315 DHPFDVKWELTAN-----NVIMVSAIGNDGLYGLTLNPNADQMDVIGVGIDFDENTAR 369
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 240 --PVGQSOTLEAVNQAATNAGVLVVAATNGNGS--GTVSYPARYANALAYGATDQNNRAS 295

QY 370 FSSRGMTTWELPGYGRMKRPDIYTYAGVRGSGVKGGRALSGTSVSPVAVGAVTLVLS 429
   | : | | | | | | | : | : | : | : | : | : | : |
Db 296 FSOYGTGL-----NIVAPCGVQSTYPCNRYASLSGTSMTATPHVAGVAALV-- 341

QY 430 TVQKRELVPASMKQALIASARRLPVGNMFEQCH 462
   | : | : | : | : | : | : | : | : | : | : | : | : |
Db 342 -KQKNPSNNTQIRHLSLSTATSLGNSNQFGSG 373

RESULT 12
Q45301
ID Q45301 PRELIMINARY; PRT; 379 AA.
AC Q45301;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```



```

Db 2 MRKSFVGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
Qy 53 EYEIVAF-NGYFTAKARNPSISSALKSSEVD--NWRIIPRNPSSDYPSPDFEVIQKEKQ 110
Db 37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRINIIRAKAKLD-----80
Qy 111 KAGLLTLEDHPIKRVTPQKVFRLSKLYAESDPTVPCNETRWSQKWSSRLRRASLSLG 170
Db 81 KEALKEVKNPDV-----AYVEED-----99
Qy 171 SGFWHATGRSSRLRAIPROVAQTLOADVLMQWGYTGANVRVAVEDTGLSEKHPFKN 230
Db 100 -----HVAHALAQTVPYGI--PLIKADKVQAGFGKANVKVAVLDTGIQASHPDLNV 149
Qy 231 VKERTNNTERTLDDGLGHGTFVAGVIASRECO---GFAPDAELHIFRVTNNQVSYTS 287
Db 150 VGGASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVPASVLYAVKVLNNSGSGSYS 209
Qy 288 WFLDAFNAILKIDVNLNLSIGGPDFMD--HPFVDKVELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIVSGIEWATTNGMDVINMNSLGGASGSGTAMKQAVDNAY---ARGVVVVAAAGNSGSGNT 266
Qy 344 GTLNAPADQMDVIGVGIGDFEDNIARESSRGMTTWELPGGYGRMKPDIVTYGAGVRSV 403
Db 267 NTIGYPKYDSVIAVAGVDSNSRASFSGA---EL-----EVMAPGAGVYSTYP 314
Qy 404 KGGCRALSGTSVSPVAVAGVTLVSTVQKRELNPASMKQALIASARRLPGVNMFEGH 463
Db 315 TNYATLNGTSMVSPHVAGAAALILS---KPNLSASQVRNRLSSTATYL--SSFYGK 369
Qy 464 GKDLLRAYQ 473
Db 370 GLINVEAAQ 379

```

RESULT 17

```

Q9F941 ID Q9F941 PRELIMINARY; PRT; 374 AA.
AC Q9F941;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Subtilisin precursor (EC 3.4.21.62) (Fragment).
GN SUBC.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_Taxid=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC12759;
RX MEDLINE=20561724; PubMed=11109488;
RA Evans K.L., Crowder J.S., Miller E.S.;
RT "Subtilisins of Bacillus spp. hydrolyze keratin and allow growth on
feathers."
RL Can. J. Microbiol. 46:1004-1011(2000).
DR EMBL; AF205191; AAG31028.1; -.
DR HSP; P00780; ICSE.
DR InterPro; IPR00209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PRO0723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Signal.
FT SIGNAL 1. 29 POTENTIAL.
FT CHAIN 30 >374 SUBTILISIN.
FT NON_TER 374 374
SQ SEQUENCE 374 AA; 38423 MW; 640142EB813D0E60 CRC64;

```

Query Match 5.68; Score 316.5; DB 2; Length 374;
 Best Local Similarity 24.9%; Pred. No. 8.5e-15;

```

Matches 121; Conservative 65; Mismatches 170; Indels 129; Gaps 17;
Qy 1 MKNLVNIWLLLVLLCGKKHLGDRLEKKSFEKAPCGCSHLTLKVEFSSTV-----V 52
Db 2 MRKSFVGLMLTAFM-----LVFTMAFSDSASAAQPAKNV 36
Qy 53 EYEIVAF-NGYFTAKARNPSISSALKSSEVD--NWRIIPRNPSSDYPSPDFEVIQKEKQ 110
Db 37 EKDYIVGFKSGVKTASVKKDIKES--GGKVDKQFRINIIRAKAKLD-----80
Qy 111 KAGLLTLEDHPIKRVTPQKVFRLSKLYAESDPTVPCNETRWSQKWSSRLRRASLSLG 170
Db 81 KEALKEVKNPDV-----AYVEED-----99
Qy 171 SGFWHATGRSSRLRAIPROVAQTLOADVLMQWGYTGANVRVAVEDTGLSEKHPFKN 230
Db 100 -----HVAHALAQTVPYGI--PLIKADKVQAGFGKANVKVAVLDTGIQASHPDLNV 149
Qy 231 VKERTNNTERTLDDGLGHGTFVAGVIASRECO---GFAPDAELHIFRVTNNQVSYTS 287
Db 150 VGGASFVAGEAYNTDGNHGHGTHVAGTVAALDNTTGVLGVPASVLYAVKVLNNSGSGSYS 209
Qy 288 WFLDAFNAILKIDVNLNLSIGGPDFMD--HPFVDKVELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIVSGIEWATTNGMDVINMNSLGGASGSGTAMKQAVDNAY---ARGVVVVAAAGNSGSGNT 266
Qy 344 GTLNAPADQMDVIGVGIGDFEDNIARESSRGMTTWELPGGYGRMKPDIVTYGAGVRSV 403
Db 267 NTIGYPKYDSVIAVAGVDSNSRASFSGA---EL-----EVMAPGAGVYSTYP 314
Qy 404 KGGCRALSGTSVSPVAVAGVTLVSTVQKRELNPASMKQALIASARRLPGVNMFEGH 463
Db 315 TNYATLNGTSMVSPHVAGAAALILS---KPNLSASQVRNRLSSTATYL--SSFYGK 369
Qy 464 GKDLL 468
Db 370 GLINV 374

RESULT 18
Q93Q20 ID Q93Q20 PRELIMINARY; PRT; 312 AA.
AC Q93Q20;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ISPD.
GN ISPD.
OS Clostridium difficile.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_Taxid=1496;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=630; TRANSPOSON-TN5398;
RX MEDLINE=20106991; PubMed=10639372;
RA Farrow K.A., Lytras D., Rood J.I.;
RT "The macrolide-lincosamide-streptogramin B resistance determinant from
Clostridium difficile 630 contains two erm(B) genes."
RL Antimicrob. Agents Chemother. 44:411-413(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=630; TRANSPOSON-TN5398;
RA Farrow K.A., Lytras D., Rood J.I.;
RT "Genomic Analysis of the Erythromycin Resistance Element Tn5398 from
Clostridium difficile."
RL Microbiology 0:0-0(2001).
DR EMBL; AF109075; AAK7652.1; -.
DR MEROPS; S08.030; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE; PS00137; SUBTILASE_HIS; UNKNOWN_1.

```

```

DR PROSITE; PS00138; SUBTILASE_SER; UNKNOWN1.
SQ SEQUENCE 312 AA; 34104 MW; 5F365313EB3DB78B CRC64;

Query Match
Best Local Similarity 28.4%; Pred. No. 8.8e-15; Length 312;
Matches 92; Conservative 60; Mismatches 119; Indels 53; Gaps 11;

QY 186 LRAIPROVAOTLQ-----ADVLQMGYTGANVRVAVFDLGLSEKHPHF-- 229
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 MKLIPYEINENLRGAKNFFYGIQKNARGMWDEYTGKNIVVIGIDTGCDSHPLKKG 64
QY 230 -----NVKERTNWTNERTLDDGLGHGTFVAGVIASM---RECQGFAPDAELHIFRVFTNN 281
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 65 IIGGANFSDSN-GNKNIYEDFNHGCHTVAGITAAASNNNEVGVAPDCKLLIAKALNKD 123
QY 282 QVSYTSMFIDAFNYAILKIDVLNLSIGGDPDMDFPVQKVMWELTANNVIMVSAIGNDGP 341
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 GTETYQSIINAFVANNKVDIISLGG-NKDDKLNKNAVQAVKNNISVVCAGNNGD 182
QY 342 LYGLTANN---PADQMDVIGGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGV 398
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 183 GDSSTSEYSPASVAEIVGAINENYLVEKFSNNTTI-----DIVAPGRNI 230
QY 399 RGSVKGCCRALSGTSTVASPVVAGAVTLVSTVQ---KRELNPASMKQALIASARELPQ 455
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 231 ISYMDNKLAMSTGMSAPYVSSGLALIKWAREEFERDL-DEALYAQLIKLCTRAL-G 288
QY 456 VNMFEQGHCKDLRLRAYQILNSYK 479
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 289 IPTREGNG-----YLYLNLTK 305

RESULT 19
Q9DFD3
ID Q9DFD3 PRELIMINARY; PRT; 310 AA.
AC Q9DFD3
DT 01-MAR-2001 (TREMELrel. 16, Created)
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Kera (Fragment).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N57N1;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282894; RAG00493.1;
DR HSP: P00780; ICSE.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILASIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
FT NON_TER 1
SQ SEQUENCE 310 AA; 31302 MW; 21E1432B9D4E7B14 CRC64;

Query Match
Best Local Similarity 5.6%; Score 313; DB 2; Length 310;
Matches 95; Conservative 49; Mismatches 129; Indels 28; Caps 9;

QY 180 HSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAVFDLGLSEKHPHF-- 239
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 HVAHALAQTPVGI-PLIKADKVOAGYKGANVRVAVLDTGIAQSHPDNLNVGGASPVAG 89
QY 240 ERTLDGLGHGTFVAGVIASMRCO---GEAPDAELHIFRVFTNNQVSTWFLDAFNVA 296
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 EAYNTDGNHGCHTVAGTVAALDNTTGVLGAPNVSLYANKVNLNSGSGSYSGVIGLEWA 149

297 ILKKIDVLNLSIGGDPDMF--HPFDKVMWELTANNVIMVSAIGNDGL--YGTLLNPADO 352
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 TTNGMDVINMNSIGGASGTAMKQAVDNAT---ARGVVVAAAGSSGSGNTWTIGYPAY 206
QY 353 MDVIGVGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVSGVKGCCRALSG 412
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 207 DSVIAYGAVDSNRSRFSVGA---EL-----EVMAPGAGVYTYPTSTVATLNG 254
QY 413 TSVASPVVAGAVTLVSTVQKRELNPASMKQALIASARRLPVGNMFEQGHCKDLRLRAY 472
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 255 TSMSAPHVAGAAALILS---KHPNLSASQVRNLSSTATYLG--SSFTYKGLINVEAAA 309
QY 473 Q 473
DQ 310 Q 310

RESULT 20
O87655
ID O87655 PRELIMINARY; PRT; 382 AA.
AC O87655
DT 01-NOV-1998 (TREMELrel. 08, Created)
DT 01-NOV-1998 (TREMELrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE Subtilisin.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K-54;
RA Yoo C.K., Kang S.M.;
RT "Characterization of nucleotide sequence for thermostable alkaline
RT protease with strong fibrinolytic activity from B. subtilis K-54.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF093112; AAC63365.1;
DR HSP: P00782; 1SPB.
DR MEROPS: S08.034; -.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILASIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
SQ SEQUENCE 382 AA; 39254 MW; 449150F3CD0A0EE3 CRC64;

Query Match
Best Local Similarity 5.6%; Score 313; DB 2; Length 382;
Matches 88; Conservative 44; Mismatches 94; Indels 38; Gaps 11;

QY 180 HSSRLRLRAIPROVAOTLQADVLQMGYTGANVRVAVFDLGLSEKHPHF-- 233
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 102 HVAQAVQSPYGVQ-IPKALHLSQGFVGSNVKAVVIDSGIDSHDPLKAVAGASWPS 160
QY 234 RTNWTNERTLDDGLGHGTFVAGVIASMR---CQGFAPDAELHIFRVFTNNQVSTWFL 290
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 ETN-----PFQDNNSHGTHVAGTVAALNNSVFLVGVAPFSASLYAKVYLGADSGQYSWII 215
QY 291 DAFNYAILKIDVLNLSIGGDPDMF--HPFDKVMWELTANNVIMVSAIGNDGLYQ--TL 346
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 216 NGIEWAIANNMDVINMNSIGGPGSAAKAAVDKA---VASGVVVAAGNEGTSQGSSTV 272
QY 347 NNPADQMDVIGGGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVSG--GVK 404
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 273 GYPGKVPYVIAVAGVNSNQRASFSSVGS---EL-----DVMAPGVSIQSTLPNGK 320
QY 405 GGCRLSGTSVASPVVAGAVTLV 428
DQ 321 YG--AYNGTSMASPHVAGAAAIL 342

```



```
Db 274 AQ 275
RESULT 23
Q9FDF2 PRELIMINARY; PRT; 310 AA.
AC Q9FDF2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Kera (Fragment).
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=837B;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282895; AAC00494.1;
DR HSP: P00780; LCSE.
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: P00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
FT NON_TER
SQ SEQUENCE 310 AA; 31285 MW; 740B9816277876EA CRC64;

Query Match 5.5%; Score 308; DB 2; Length 310;
Best Local Similarity 31.2%; Pred. No. 2.6e-14;
Matches 94; Conservative 49; Mismatches 130; Indels 28; Gaps 9;

QY 180 HSRRLRAIPROVAQTLOADVLMQMGYTGANVRVAVFDTGLSKHPHFKN-VKERTNNMTN 239
Db 31 HVAHALAQTPYGI-PLIKADKVOAQGFKANVKAVLDTGQASHDPLNVVGASFAVG 89
QY 240 ERTLDGGLGHGTFVAGVIASMRQ---GFAPDAELHIFRVTNNVSYTSWFLDAFNYA 296
Db 90 EAYNTDGNHGHTVAGTVAALDNTTGLGVAPSVSLYAVKLVNSGSGSYGVSGIEMA 149
QY 297 ILKIDVNLISIGGPDVMD--HPFVDKWLVTANNVIMVSAIGNDGPL--YGTLLNPADQ 352
Db 150 TTNGMDVNLISLGSGSSTANKQAVDNAY--ARGVVVVAAGNSGSGNTNIGYPKY 206
QY 353 MDVIGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGVSGVKGCCRALSG 412
Db 207 DSVIANGVDSNRSFSSVGA--EL-----EVMAPGAGYSTPTPTATYATNG 254
QY 413 TSVASPVVAGAVTLVSTVOKRELVPASMKQALIASARRLPGVNMFEQGHGKLDLLRAY 472
Db 255 TSMSAPHVAGAAALILS---RHPNLASQVRNRLSATYLG--SSFYKGLINVEAAA 309
QY 473 Q 473
Db 310 Q 310

RESULT 24
Q45621 PRELIMINARY; PRT; 323 AA.
AC Q45621;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE A intracellular serine protease.
GN ISPO.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
```

```
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NKS-21;
RA Yamagata Y., Ichishima E.;
RT "A new alkaline serine protease from alkalophilic bacillus sp.:
RT cloning, sequencing, and characterization of an intracellular
RT protease."
RL Curr. Microbiol. 30:357-366(1995).
DR EMBL: D37921; BAA07142.1;
DR HSP: Q99405; 1MPT.
DR MEROPS: S08.030;
DR InterPro: IPR00209; Peptidase_S8.
DR Pfam: P00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KW Protease.
SQ SEQUENCE 323 AA; 34643 MW; 6D20F71A73F1AD96 CRC64;

Query Match 5.4%; Score 304.5; DB 2; Length 323;
Best Local Similarity 28.8%; Pred. No. 5.1e-14;
Matches 91; Conservative 58; Mismatches 104; Indels 63; Gaps 14;

QY 187 RAIPROVAQTLOADVLMQMGYTGANVRVAVFDTGLSKHPHFKN-VKERTNNMTN 244
Db 19 KVIPPGI-EMIEAPAVWEAGYKGGNTVAVLDTGCTTHIEFKDQIIDGRNFTDSDP 77
QY 245 ---DLGLGHGTFVAGVIASMRQ---GFAPDAELHIFRVTNNVSYTSWFLDAFN 294
Db 78 DVEDSNGHGHVCGPVAA---CENDKGVIGTAPKAKLVVKSQGYGDTKWIIEGVR 134
QY 295 YAI-----LKKIDVNLISIGGPDVMD--HPFVDKWLVTANNVIMVSAIGNDGPLYGLN 347
Db 135 YAINVRGPNRVRVISMISLGR--IDTPELQAIKHAVAEDILVVCAAGNE---GDGN 188
QY 348 NPADQM-----DVIGVGIDFEDNIARFSSRGMTTWELPGYGRMKPDIVTYGAGVRG 400
Db 189 HDTDEYAVPGAYPEVQVGSVNLEGEISRFSTNCAL-----DLVAPGEEIIS 236
QY 401 SGVKGCCRALSGTSVASPVVAGAVTLVSTVOK---RELVPASMKQAL-----IASARR 452
Db 237 TYLNGYAVLSGTSMTATPHVSGAALLIEQVEKEFERKLEPEIPAQLIKHTVSLNFSRR 296
QY 453 LFGVNMFEQGHGKLDL 468
Db 297 -----AQSGLLKL 305

RESULT 25
Q9F7C2 PRELIMINARY; PRT; 310 AA.
AC Q9F7C2;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Keratinolytic protease Kera (Fragment).
GN KERA.
OS Bacillus licheniformis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1402;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OWU1432B;
RA Goldstein G., Wagner E.K., Lee Y.E., Masisak E.S., Oh J.;
RT "Nucleotide sequence of kera gene encoding a keratinolytic protease of
RT Bacillus licheniformis OWU 1432B."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF282895; AAC10033.1;
```

DR HSP; P00780; ICSE.
DR MEROPS; S08.001; -.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
KW Protease.
FT NON_TER
SQ SEQUENCE 1 310 AA; 31271 MW; 420B981623EAD83 CRC64;
Query Match 5.4%; Score 304; DB 2; Length 310;
Best Local Similarity 30.9%; Pred. No. 5.1e-14;
Matches 93; Conservative 49; Mismatches 131; Indels 28; Gaps 9;
QY 180 HSSRLRLRAIPROVAQTLOADVLWOMCYTCANRVAVFDGLSEKHPFKNVKERTNWTN 239
DB 31 HVAHALAQVYPI-PLIRADKYQAOGFKGANVKVAVLDGTIOASHPDNLNVGGASFVAG 89
QY 240 ERTLDGDLGHGTFVAGVIASMRQC---GFAPDAELHIFRVFTNNQVSYTSWFLDAFNVA 296
DB 90 EAYNTDGNHGTHVAGTVAALDNTTGVLGVPSPVSLYAVKVLNSSGSGSYSGIVSGIEWA 149
QY 297 ILKKIDVNLISGGPDFMD--HPFDKVKWELTANNVIMVSAIGNDGL--YGTINNPADQ 352
DB 150 TTNGMDVINNSLGGSGSTAMKQAVDNAY---ARGVVVAAAGNSGSGSGNTNTIGYPAKY 206
QY 353 MDVIGVGGIDFEDNIARESSRGMTTWELPGYGRMKPDIVTYGAGYRGSGVKGCCRALSG 412
DB 207 DSVIAGVADSNRASFSSVGA---EL-----EVNAPGAGVYSTYPTNTYATLNG 254
QY 413 TSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARLPGVNMFEQGHGKLDLLRAY 472
DB 255 TSMGSPHVAGAAALILS---KHPNLSASQVRNLSSTATYLG--SSFYIGKGLINVEAAA 309
QY 473 Q 473
DB 310 Q 310

Search completed: May 29, 2003, 13:38:23
Job time : 106 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:28:27 ; Search time 25 seconds
(without alignments)
1745.324 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIVLLVLLCGKKH.....PRVKRQLMQVHPKPTSV 1052

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5617	100.0	1052	1	MSIP_HUMAN
2	5487	97.7	1052	1	MSIP_CRIGR
3	5476	97.5	1052	1	MSIP_RAT
4	5477	96.6	1052	1	MSIP_MOUSE
5	340	6.1	381	1	SUBT_BACST
6	339	6.0	381	1	SUBT_BACSU
7	338.5	6.0	381	1	SUBN_BACNA
8	338.5	6.0	381	1	SUBT_BACSA
9	329.5	5.9	379	1	SUBT_BACLI
10	319.5	5.7	382	1	SUBT_BACAM
11	316.5	5.6	275	1	SUBT_BACPU
12	308	5.5	274	1	SUBD_BACLI
13	299	5.3	326	1	ISP_PAEPO
14	291	5.2	319	1	ISPL_BACSU
15	285	5.1	1398	1	PLS_PYRFU
16	284.5	5.1	321	1	ISP_BACCS
17	283	5.0	380	1	ELVA_BACAO
18	282	5.0	380	1	ELVA_BACCS
19	267.5	4.8	269	1	SUBS_BACLE
20	266	4.7	378	1	ELVA_BACSP
21	262.5	4.7	269	1	PRTM_BACSP
22	262.5	4.7	682	1	NISP_LACLA
23	259.5	4.6	269	1	SUBB_BACLE
24	249	4.4	806	1	SUBV_BACSU
25	243	4.3	645	1	SUBE_BACSU
26	241.5	4.3	420	1	SUBT_BACSU
27	240	4.3	401	1	THES_BACSU
28	230.5	4.1	361	1	ELVA_BACPD
29	225.5	4.0	1433	1	SUBF_BACSU
30	223	4.0	279	1	THET_THEVU
31	220.5	3.9	595	1	BPRX_BACSU
32	220.5	3.9	894	1	WPRA_BACSU
33	219	3.9	422	1	TKSU_PYRKO

34 215.5 3.8 461 1 EPIP_STAEP
35 215.5 3.8 1374 1 YQ56_CABEL
36 195.5 3.5 513 1 AOL1_THEAQ
37 195 3.5 384 1 PRFK_TRIAL
38 194 3.5 404 1 SMPL_MAGPO
39 193 3.4 635 1 PRTE_YEAST
40 191 3.4 1249 1 TPP2_HUMAN
41 190.5 3.4 1262 1 TPP2_MOUSE
42 190 3.4 1249 1 TPP2_RAT
43 189.5 3.4 1167 1 SCAL_STRPY
44 187.5 3.3 409 1 ALP_TRIHA
45 184.5 3.3 478 1 YSP3_YEAST

ALIGNMENTS

RESULT 1
MSIP_HUMAN
ID MSIP_HUMAN STANDARD; PRT; 1052 AA.
AC Q14703.02UE57;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE ~~Membrane-bound~~ transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin-isozyme-1) (SKI-1).
GN MBTPS1 OR SIP OR SKI1 OR KIAA0091.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid;
RX MEDLINE=20401147; PubMed=10944850;
RA Nakajima T., Iwaki K., Kodama T., Inazawa J., Emi M.;
RT "Genomic structure and chromosomal mapping of the human site-1
RT protease (SIP) gene.";
RL J. Hum. Genet. 45:212-217(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=95308325; PubMed=7788527;
RA Nagase T., Miyajima N., Tanaka A., Sazuka T., Seki N., Sato S.,
RA Tabata S., Ishikawa K.-I., Kawabayashi Y., Kotani H., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. III.
RT The coding sequences of 40 new genes (KIAA0081-KIAA0120) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL PWA Res. 2:37-43(1995).
RN [3]
RP SEQUENCE OF 735-1052 FROM N.A.
RC TISSUE=Testis;
RA Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION, PARTIAL SEQUENCE, AND MASS SPECTROMETRY.
RX MEDLINE=20112790; PubMed=10644685;
RA Toure B.B., Munzer J.S., Basak A., Benjannet S., Rochemont J.,
RA Lazure C., Chretien M., Seidah N.G.;
RT "Biosynthesis and enzymatic characterization of human SKI-1/SIP and
RT the processing of its inhibitory prosegment.";
RL J. Biol. Chem. 275:2349-2358(2000).
RN [5]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99145548; PubMed=9990022;
RA Seidah N.G., Mowla S.J., Hamelin J., Mamarbachi A.M., Benjannet S.,
RA Toure B.B., Basak A., Munzer J.S., Marcinkiewicz J., Zhong M.,
RA Barale J.-C., Lazure C., Murphy R.A., Chretien M., Marcinkiewicz M.;
RT "Mammalian subtilisin/kexin isozyne SKI-1: A widely expressed
RT proprotein convertase with a unique cleavage specificity and cellular
RT localization.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1321-1326(1999).
CC -!- FUNCTION: Catalyzes the first step in the proteolytic activation

ID MSIP_CRIGR STANDARD; PRT; 1052 AA.
AC Q922A8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-bound transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin-isozyme-1) (SKI-1)
GN MBTPS1 OR SIP OR SKI1.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF ASP-218; HIS-249 AND SER-414.
RC STRAIN=25-RA;
RX MEDLINE=99026600; PubMed=9809072;
RA Sakai J., Rawson R.B., Espenshade P.J., Cheng D., Seegmiller A.C.,
RA Goldstein J.L., Brown M.S.;
RT "Molecular identification of the sterol-regulated luminal protease
RT that cleaves SREBPs and controls lipid composition of animal cells.";
RL Mol. Cell 2:505-514(1998).
RN [2]
RP FUNCTION.
RX MEDLINE=21111045; PubMed=11163209;
RA Ye J., Rawson R.B., Komuro R., Chen X., Dave U.P., Prywes R.,
RA Brown M.S., Goldstein J.L.;
RT "ER stress induces cleavage of membrane-bound ATF6 by the same
RT proteases that process SREBPs.";
RL Mol. Cell 6:1335-1344(2000).
CC FUNCTION: Catalyzes the first step in the proteolytic activation
CC of the sterol regulatory element-binding proteins (SREBPs).
CC Other known substrates are BDNF and ATF6.
CC -1- CATALYTIC ACTIVITY: Cleaves after hydrophobic or small residues,
CC provided that Arg or Lys is in position P4. Cleaves known
CC substrates after Arg-Ser-Val-Leu (SERBP-2), Arg-His-Leu-Leu
CC (ATF6), Arg-Gly-Leu-Thr (BDNF) and its own propeptide after Arg-
CC Arg-Leu-Leu.
CC -1- COFACTOR: Calcium-dependent (By similarity).
CC -1- ENZYME REGULATION: Inhibited by divalent copper and zinc ions, but
CC not by nickel or cobalt. Inhibited by its prosegment, but not
CC smaller fragments thereof (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum and Golgi. May sort to other organelles, including
CC lysosomal and/or endosomal compartments (By similarity).
CC -1- INDUCTION: Down-regulated by sterols.
CC -1- PTM: The 148 kDa zymogen is processed progressively into two
CC membrane-bound 120 and 106 kDa forms in the endoplasmic reticulum,
CC and late into a secreted 98 kDa form. The propeptide is
CC autocatalytically removed through an intramolecular cleavage after
CC Leu-186. Further cleavage generates 14, 10, and 8kDa
CC intermediates (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: AF078105; AAC78321.1;
CC HSP: Q99405; 1MPT.
CC MEROPS: S08.063;
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS: PR00723; SUBTILISIN.
CC PROSITE: PS00136; SUBTILASE.ASP; FALSE_NEG.
CC PROSITE: PS00137; SUBTILASE.HIS; 1.
CC PROSITE: PS00138; SUBTILASE.SER; 1.
CC Hydrolase; Protease; Serine protease; Lipid metabolism;
KW

Cholesterol metabolism; Signal; Transmembrane; Endoplasmic reticulum;
Golgi stack; Zymogen; Autocatalytic cleavage; Glycoprotein; Calcium.
KW SIGNAL 1 17
KW PROPEP 18 186
KW CHAIN 187 1052
FT POTENTIAL.
FT POTENTIAL.
FT MEMBRANE-BOUND TRANSCRIPTION FACTOR
FT SITE-1 PROTEASE.
FT LUMENAL (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT SERINE PROTEASE.
FT ARG/LYS/PRO-RICH (BASIC).
FT CLEAVAGE (AUTO-) (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CHARGE RELAY SYSTEM (BY SIMILARITY).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT D-N: LOSS OF ACTIVITY.
FT H->P: LOSS OF ACTIVITY.
FT S->A: LOSS OF ACTIVITY.
SQ SEQUENCE 1052 AA; 117564 MW; D/DBAFE988B4C69C CRC64;
Query Match Score 5487; DB 1; Length 1052;
Best Local Similarity 97.7%;
Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;
QY 1 MKLVNIWLLLVLLCGKHLGDRLEKKSEKAPCGCSHLTLKVEFSSTVVEYEVAF 60
DB 1 MKLVNIWLLLVLLCGKHLGDRLEKKSEKAPCGCSHLTLKVEFSSTVVEYEVAF 60
QY 61 NGYFTAKARNSFISSALKSEVDNWRILPNNSSDYPDSFVQIQLKQAGALLTLEDH 120
DB 61 NGYFTAKARNSFISSALKSEVDNWRILPNNSSDYPDSFVQIQLKQAGALLTLEDH 120
QY 121 PNKRVTPQKRVFSLKFAESDPTVPCNETRWSQKQSSRPLRRASISLGSFGFWHATGRH 180
DB 121 PNKRVTPQKRVFSLKFAESDPTVPCNETRWSQKQSSRPLRRASISLGSFGFWHATGRH 180
QY 181 SSRLRLRAIPRQVAQTLOADVLMQMGYTGANRVAVEDTGLSEKHPKFNKERTNWTNE 240
DB 181 SSRLRLRAIPRQVAQTLOADVLMQMGYTGANRVAVEDTGLSEKHPKFNKERTNWTNE 240
QY 241 RTLDDGLGHGTFVAGVIASMRCEGFPADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
DB 241 RTLDDGLGHGTFVAGVIASMRCEGFPADAEHLHFRVFTNNQVSYTSWFLDAFNAILKK 300
QY 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGTLANPADQMDVIGVG 360
DB 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVIMVSAIGNDGLYGTLANPADQMDVIGVG 360
QY 361 IDVEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRSGVKGCCALSGTSVSPV 420
DB 361 IDVEDNIAFSSRGMTTWELPGGYGRMKPDIVTYGAGVGRSGVKGCCALSGTSVSPV 420
QY 421 AGAVTLVSVQRELNVNPMQALIASARLLPGVNMFGCHGKLDLLRAYQILNYSKP 480
DB 421 AGAVTLVSVQRELNVNPMQALIASARLLPGVNMFGCHGKLDLLRAYQILNYSKP 480
QY 481 QASLSPSYIDLTCPYMWPCYCSPIYYGGMPTVNVNVTILNMGVTVGRVDPKQMPYLPQ 540
DB 481 QASLSPSYIDLTCPYMWPCYCSPIYYGGMPTVNVNVTILNMGVTVGRVDPKQMPYLPQ 540
QY 541 NGDNIEVAFSYSSVLPWPSGYLAISISVTKKAASWEGIAQGHVIMTVASPAETESKNGAE 600
DB 541 NGDNIEVAFSYSSVLPWPSGYLAISISVTKKAASWEGIAQGHVIMTVASPAETESKNGAE 600
QY 601 QTSTVKLPKIKVKIIPTPPRSKRVLDQYHNLRTPPGYPFRDNLBMKNDPLDWDGHDHHTN 660
DB 601 HTSTVKLPKIKVKIIPTPPRSKRVLDQYHNLRTPPGYPFRDNLBMKNDPLDWDGHDHHTN 660

Db 241 RTLDGGLGHGTFVAGVATASRECGFADPAELHFRFTNNQVSYTWFDAFYAIKK 300
Qy 301 IDVLNLSTGGDFDMDFVVKWELTANNVIMWSAIGNDGLYGLTANPADQMDVIGVG 360
Db 301 MDVLNLSTGGDFDMDFVVKWELTANNVIMWSAIGNDGLYGLTANPADQMDVIGVG 360
Qy 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGKGCALSGTSSVSPV 420
Db 361 IDFDNTARSSRGMTTWELPGGYGRMKPDIVTYGAGVGRGKGCALSGTSSVSPV 420
Qy 421 AGAVTLVSVTKRELNPASMKQALIASARRLPVGNVMEGCGKLDLLRAYQILNSYK 480
Db 421 AGAVTLVSVTKRELNPASVQKALIASARRLPVGNVMEGCGKLDLLRAYQILNSYK 480
Qy 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMTVVNVTILNMGVGTGRIVDKPDQVPLPQ 540
Db 481 QASLSPSYIDLTECPYMWPCSQPIYYGGMTVVNVTILNMGVGTGRIVDKPDQVPLPQ 540
Qy 541 NGDNIEVAFSSVSLVPMWSGLAISISVTKKAASWEGIAQGHVIMITVASPAETESKNGAE 600
Db 541 NGDNIEVAFSSVSLVPMWSGLAISISVTKKAASWEGIAQGHVIMITVASPAETESKNGAE 600
Qy 601 QTSVTKLPIKVIPTPPRSKRVLDQYHNLRTPPGFPDRNLRMKNNDPLDNGDHTN 660
Db 601 HTSTVKLPIKVIPTPPRSKRVLDQYHNLRTPPGFPDRNLRMKNNDPLDNGDHTN 660
Qy 661 FRDMYQHLRSMGVFVEVLGAPFTCFDASQYGTLLMDSEEEYPEETAKLRDNDGLSL 720
Db 661 FRDMYQHLRSMGVFVEVLGAPFTCFDQYGTLLMDSEEEYPEETAKLRDNDGLSL 720
Qy 721 VIFSDWNTSVMRKVFYDENTQWMPDPTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
Db 721 VIFSDWNTSVMRKVFYDENTQWMPDPTGGANIPALNELLSSVWNGFSDGLYEGEFTL 780
Qy 781 ANHDMYASCSGTAKEPEDGVITQTFKDGLEVLKQETAVENVPILGLYQIYPAEGGR 840
Db 781 ANHDMYASCSGTAKEPEDGVITQTFKDGLEVLKQETAVENVPILGLYQIYPAEGGR 840
Qy 841 IVLYGDSNCLDSSHQKDCFWLLDALQYTSYGVTPSLSHSGNRQRPSPGAGSVTPERM 900
Db 841 IVLYGDSNCLDSSHQKDCFWLLDALQYTSYGVTPSLSHSGNRQRPSPGAGSVTPERM 900
Qy 901 EGNHLHRSKVLBAHLGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLISIDLKV 960
Db 901 EGNHLHRSKVLBAHLGDKPRPLPACPRLSWAKPOPLNETAPSNLWKHKLISIDLKV 960
Qy 961 VLPNFRSNRQVRLSPGSGGANDIPGIMPGRYNQEVGOTIPVFAFLGAMVLAFFVQ 1020
Db 961 VLPNFRSNRQVRLSPGSGGANDIPGIMPGRYNQEVGOTIPVFAFLGAMVLAFFVQ 1020
Qy 1021 INKAKSRPKRRKPRVKRQPLMQCVHPKPTSV 1052
Db 1021 ISKAKSRPKRRKPRVKRQPLMQCVHPKPTSV 1052

RESULT 4

MS1P_MOUSE ID MS1P_MOUSE STANDARD; PRT; 1052 AA.
AC ORNT2.
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Membrane-bound transcription factor site-1 protease precursor
DE (EC 3.4.21.-) (Site-1 protease) (Subtilisin/kexin isozyme-1) (SKI-1)
DE (Sterol-regulated luminal protease).
GN MRPSP1 OR SIP OR SKI1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RC TISSUE=Pituitary;
RX MEDLINE=99145548; PubMed=9990022;
RA Seidah N.G., Mowla S.J., Hamelin J., Mamabachi A.M., Benjannet S.,
RA Toure B.B., Basak A., Munzer J.S., Marcinkiewicz J., Zhong M.,
RA Barale J.C., Izure C., Murphy R.A., Chretien M., Marcinkiewicz M.,
RT "Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
RT proprotein convertase with a unique cleavage specificity and cellular
RT localization.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:1321-1326(1999).
RN
RP
SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Catalyzes the first step in the proteolytic activation
CC of the sterol regulatory element-binding proteins (SREBPs).
CC Other known substrates are BDNF and ATF6 (By similarity).
CC -1- CATALYTIC ACTIVITY: Cleaves after hydrophobic or small residues,
CC provided that Arg or Lys is in position P4. Cleaves known
CC substrates after Arg-Ser-Val-Leu (SERBP-2), Arg-His-Leu-Leu
CC (ATF6), Arg-Gly-Leu-Thr (BDNF) and its own propeptide after Arg-
CC Arg-Leu-Leu.
CC -1- COFACTOR: Calcium-dependent (By similarity).
CC -1- ENZYME REGULATION: Inhibited by divalent copper and zinc ions, but
CC not by nickel or cobalt. Inhibited by its prosegment, but not
CC smaller fragments thereof (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein; endoplasmic
CC reticulum and Golgi. May sort to other organelles, including
CC lysosomal and/or endosomal compartments (By similarity).
CC -1- PFM: The 148 kDa zymogen is processed progressively into two
CC membrane-bound 120 and 106 kDa forms in the endoplasmic reticulum,
CC and late into a secreted 98 kDa form. The propeptide is
CC autocatalytically removed through an intramolecular cleavage after
CC Leu-186. Further cleavage generates 14, 10, and 8kDa
CC intermediates (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF094820; A027010.1;
CC EMBL; BC011533; A011533.1;
CC HSP; Q99405; IMPT.
CC MEROPS; S08.063;
CC MGD; MGI:1927235; Mbtps1.
CC InterPro: IPR000209; Peptidase_S8.
CC Pfam: PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Protease; Serine protease; Lipid metabolism;
CC Cholesterol metabolism; Signal; Transmembrane; Endoplasmic reticulum;
CC Golgi stack; Zymogen; Autocatalytic cleavage; Glycoprotein; Calcium.
CC SIGNAL 1 17
CC PROPEP 18 186
CC CHAIN 187 1052
CC
CC DOMAIN 187 999
CC TRANSMEM 1000 1022
CC CYTOPLASMIC (POTENTIAL).
CC SERINE PROTEASE.
CC DOMAIN 218 414
CC DOMAIN 1023 1050
CC ARG/LYS/PRO-RICH (BASIC).
CC SITE 186 187
CC CLEAVAGE (AUTO-) (BY SIMILARITY).
CC ACT_SITE 218 218
CC CHARGE RELAY SYSTEM (BY SIMILARITY).
CC ACT_SITE 249 249
CC ACT_SITE 414 414
CC CARBOHYD 148 148
CC N-LINKED (GLCNAC...) (POTENTIAL).

Query Match
Best Local Similarity 95.8%; Pred. No. 0;
Matches 1008; Conservative 25; Mismatches 19; Indels 0; Gaps 0;

QY 1 MKLVNIIWLLVLLCGKRLHGLDRLEKSEKAPCPGCSHLTLKVFESSTVEYEVAF 60
DB 1 MKLVNIIWLLVLLCGKRLHGLDRLEKSEKAPCPGCSHLTLKVFESSTVEYEVAF 60

QY 61 NGYFTAKARNSEFISSALKSEVDNRIIPRNPNSSDPSDFEVIQIKKAGLLTLEH 120
DB 61 NGYFTAKARNSEFISSALKSEVDNRIIPRNPNSSDPSDFEVIQIKKAGLLTLEH 120

QY 121 PNTRKVTQKRVFSLKYAFSDPTVPCNETWQSKQWQSRPLRRASLSLGSFWHATGRH 180
DB 121 PNTRKVTQKRVFSLKYAFSDPTVPCNETWQSKQWQSRPLRRASLSLGSFWHATGRH 180

QY 181 SSRLLRAIPROVAOTLQADVLQMGYTGANRVAVFDGLSEKHPKKNVKTWTNE 240
DB 181 SSRLLRAIPROVAOTLQADVLQMGYTGANRVAVFDGLSEKHPKKNVKTWTNE 240

QY 241 RLDDGLGHGTFVAGYASRECOGFADELHIFRVFTNNQVSTWFLDAFYAILKK 300
DB 241 RLDDGLGHGTFVAGYASRECOGFADELHIFRVFTNNQVSTWFLDAFYAILKK 300

QY 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVJWSAIGNDGPLYGLTNLPADQMDVIGVG 360
DB 301 IDVLNLSIGGPDFMDHPFVDKVMWELTANNVJWSAIGNDGPLYGLTNLPADQMDVIGVG 360

QY 361 IDVDNIARFSSGTMWELPGYGRVAPDITVYAGVSGVKGCCRALSGTSVSPV 420
DB 361 IDVDNIARFSSGTMWELPGYGRVAPDITVYAGVSGVKGCCRALSGTSVSPV 420

QY 421 AGAVTLLVSTVQKRELNVASVKQALIASARLPFGVNMFEQGHGKLLRAYQILNSYK 480
DB 421 AGAVTLLVSTVQKRELNVASVKQALIASARLPFGVNMFEQGHGKLLRAYQILNSYK 480

QY 481 QASLSPSIDLTCPCYMWPCYQPIYVGGMTVNTVILNGMGTGRVYDKPWPQYLPQ 540
DB 481 QASLSPSIDLTCPCYMWPCYQPIYVGGMTVNTVILNGMGTGRVYDKPWPQYLPQ 540

QY 541 NGDTEVAFSYSSVLPWPSGYLAISVTKKASWEGIAQGHVMTVWASPAETESKNGAE 600
DB 541 NGDTEVAFSYSSVLPWPSGYLAISVTKKASWEGIAQGHVMTVWASPAETESKNGAE 600

QY 601 QTSVKLPIKVKIIPPPSKRVLMQDQHNLRYPGYPFRDNLKMKNDPLDNGDHIHTN 660
DB 601 QTSVKLPIKVKIIPPPSKRVLMQDQHNLRYPGYPFRDNLKMKNDPLDNGDHIHTN 660

QY 661 FRDMDYHLSRNGYFEVLGAPTCFDAQSQXGTLLMWYDSEEEYFPEETAKLRDNDVGLSL 720
DB 661 FRDMDYHLSRNGYFEVLGAPTCFDAQSQXGTLLMWYDSEEEYFPEETAKLRDNDVGLSL 720

QY 721 VFSDWNTSVMRKVFYDENTROMMDTGGANIPALNELLVSNMGMFSDGLYECEFTL 780
DB 721 VFSDWNTSVMRKVFYDENTROMMDTGGANIPALNELLVSNMGMFSDGLYECEFTL 780

QY 781 ANHDYASGCSIAKPPEDGVVITOTFRDQGLEVLQETAVVENVPIGLYQYPAESGG 840
DB 781 ANHDYASGCSIAKPPEDGVVITOTFRDQGLEVLQETAVVENVPIGLYQYPAESGG 840

QY 841 IVLYGDSNCLDSDHQKDFWLLDALQYTSYGYTPPSLSHSGNQRPPSGAGSVTPERM 900
DB 841 IVLYGDSNCLDSDHQKDFWLLDALQYTSYGYTPPSLSHSGNQRPPSGAGSVTPERM 900

QY 901 EGNHLHRSKYLEAHGLDPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDRV 960

DB 901 EGNHLHRSKYLEAHGLDPRPLPACPRLSWAKPOPLNETAPSNLWKHOKLLSIDLDRV 960

QY 961 VLPNFRSRRPQVRPLSPGSGAWDIPGGIMPGRYNOEVGQTPVFAFLGAMVYVAFVYQ 1020

DB 961 VLPNFRSRRPQVRPLSPGSGAWDIPGGIMPGRYNOEVGQTPVFAFLGAMVYVAFVYQ 1020

QY 1021 INKAKSRKRRKRPVKRQPMQVQHPKTPSV 1052

DB 1021 ISKAKSRKRRKRPVKRQPMQVQHPKTPSV 1052

RESULT 5

ID SUBT_BACST STANDARD; PRT; 381 AA.

AC P29142:

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Subtilisin J precursor (EC 3.4.21.62).

GN AP01

OS Bacillus stearothermophilus.

OC Bacteria; Firmicutes; Bacillales; Geobacillus.

OX NCBI_taxid=1422;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCIMB 10278 / KCTC 1823;

KA MEDLINE=9231938; PubMed=1567435;

RA "Molecular cloning of a subtilisin J gene from Bacillus stearothermophilus and its expression in Bacillus subtilis.";

RT Biochem. Biophys. Res. Commun. 184:277-282(1992).

RL IT FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE, IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.

CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity for peptide bonds, and a preference for a large uncharged residue in pl. Hydrolyzes peptide amides.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN IS NOT NECESSARY FOR NORMAL SPOULATION.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M64743; AAA22247.1; -

DR PIR; JQ1487; JQ1487.

DR HSP; P04189; LSCJ.

DR MEROPS; S08.035; -

DR InterPro; IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS; PR00723; SUBTILISIN.

DR PROSITE; PS00136; SUBTILASE_ASP; 1.

DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.

KW Hydrolase; Sporulation; Serine protease; Zymogen; Signal.

FT SIGNAL 1 29 POTENTIAL.

FT PROPEP 3 106 POTENTIAL.

FT CHAIN 107 381 SUBTILISIN J.

FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).

SQ SEQUENCE 381 AA; 39495 MW; C7A596F7629087D5 CRC64;

Query Match 6.1%; Score 340; DB 1; Length 381;
Best Local Similarity 25.1%; Pred. No. 6.3e-17;

STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPOULATION.

-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: K01988; AAA22742.1; -
EMBL: Y14083; CAA74536.1; -
EMBL: Y95109; CAB12870.1; -
EMBL: K01443; AAA22814.1; -
EMBL: M16639; AAA22744.1; -
EMBL: M31060; AAA22246.1; -
EMBL: M31925; AAA22245.1; -
PIR: A00372; SUBSI.
PDB: 1SCJ; 13-JAN-99.

MEMOPS; S08.036; -
Subtilisin; HG10190; aprE.
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.
PRINTS; PR00723; SUBTILISIN.
PROSITE; PS00136; SUBTILASE_ASP; 1.
PROSITE; PS00137; SUBTILASE_HIS; 1.
PROSITE; PS00138; SUBTILASE_SER; 1.
Hydrolase; Spolulation; Serine protease; Zymogen; Signal;
3D-structure; Complete proteome.

FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN E.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM.
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM.
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM.
FT CONFLICT 27 27 V -> A (IN REF. 1, 4 AND 5).
SQ SEQUENCE 381 AA; 39523 MW; B7D2B38562C087D5 CRC64;

Query Match 6.0%; Score 339; DB 1; Length 381;
Best Local Similarity 25.1%; Pred. No. 7.4e-17;
Matches 123; Conservative 71; Mismatches 169; Indels 128; Gaps 18;

QY 1 MKLVNILLVLLVLLCGKHGDRLEKSKPEKAPCCSHLTKAYESSIVV-----52
DB 1 MRSKLMISLFLT-----LIFTAFNSMVSQAACKSST 35

QY 53 EYEVIVANGFTAKARNSFISALKSEVDNWRILPRNPPSSDVPDFVIOKEKQA 112
DB 36 EKRYIVGF-----KQTSAMSAAKKDVI-----SEKG 63

QY 113 GLTLLE-DHPNRIKRVTPQRKVFRLSKYAESDTPVPCNETRWQSKWSSRPLRASLSGS 171
DB 64 GKVKQKRYVNAATAATLDEKAVKELK---KDPVSA-----95

QY 172 GFWHATGRHSRRLLRAIPROVAQTLQADVLQMGVGTGANRVAVFDTGLSEKHPFKNV 231
DB 96 ---YVEEDHIAHEVAQSPVPGISQ- IKAPALHSGQVGTGSNVKVAVIDSGIDSSHPLD-NV 150

QY 232 KERTNWTNERT--LDGCLGHTFVAGVIA\$MRE---CQGFAPDRLHIERFVNQVST 286
DB 151 RCGASFVPSFETPNYQDSSSHGTHVAGTIALNNSICVLGVSPSASLYAVKVLDSGQY 210

QY 287 SWFLDAFNALTKIDVNLNLSIGGDFEMD--HPFYDKVWELTANNVIMVAIGNDGP--L 342
DB 211 SWINGIEWALSNNMDVINNSLGGFTG\$TALKIVDKA---VSSGIVVAAAAGNESSGS 267

QY 343 YGTLNAPADQMDVIGGIDFEDNIAFSSRGMTTWELPGYGRMKPDIVTYGAGVRSG 402
DB 268 T\$VGYPAKPT\$TIAVGAVN\$SNQAF\$SAGS---EL-----DWAPGV\$IQ\$TL 315

QY 403 VKGCRALSGTSV\$NPWAGAVTLL\$TVOKREL\$VNP\$KQALIASARLPGVNNFEG 462

DB 316 PGCTYGANGTSMATHPVAGAAALILS---RHPTWNAQVRDRLESTATYLG--NSFYTG 370
QY 463 HGKLDLLRAYQ 473
DB 371 KGLINVOAAQ 381

RESULT 7

SUBN_BACNA STANDARD; PRT; 381 AA.

AC P35835;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Subtilisin NAT precursor (EC 3.4.21.62).
GN APRN.
OS Bacillus subtilis var. natto.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86029;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NC2-1;
RX MEDLINE=93113095; PubMed=1369081;
RA Nakamura T., Yamagata Y., Ichishima E.;
RT "Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus
subtilis (natto).";
RL Biosci. Biotechnol. Biochem. 56:1869-1871(1992).
CC -!- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
for peptide bonds, and a preference for a large uncharged residue
in pI. Hydrolyzes peptide amides.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; D253319; BAA04989.1; -
EMBL; S51909; AAC60424.1; -
PIR; JH0778; JH0778.
HSSP; P07518; IMEE.
MEROPS; S08.044; -
InterPro; IPR000209; Peptidase_S8.
Pfam; PF00082; Peptidase_S8; 1.
PRINTS; PR00723; SUBTILISIN.
PROSITE; PS00136; SUBTILASE_ASP; 1.
PROSITE; PS00137; SUBTILASE_HIS; 1.
PROSITE; PS00138; SUBTILASE_SER; 1.
Hydrolase; Spolulation; Serine protease; Zymogen; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 106 POTENTIAL.
FT CHAIN 107 381 SUBTILISIN NAT.
FT ACT_SITE 138 138 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 327 327 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 381 AA; 39507 MW; DAED4B16ED1BA092 CRC64;

Query Match 5.0%; Score 338.5; DB 1; Length 381;
Best Local Similarity 27.1%; Pred. No. 8e-17;
Matches 118; Conservative 65; Mismatches 156; Indels 97; Gaps 17;

QY 48 SSTVVEYIVAFNGVFTAKARNSFISALKSEVDNWRILPRNPPSSDVPDFVIOIK 107

FT HELIX 374 377
FT TURN 378 378
SQ SEQUENCE 379 AA; 38908 MW; F19A6DC5761FB504 CRC64;

Query Match 5.9%; Score 329.5; DB 1; Length 379;
Best Local Similarity 25.3%; Pred. No. 3.5e-16;
Matches 124; Conservative 66; Mismatches 171; Gaps 17;

QY 1 MKLVNTWLLLLVLCCKHGLDRLEKSKFEKAPCPGCSHLTLKVFSTV-----V 52
DB 2 MRKSWLGLMTAFM-----LVTMAFSDSASAAQPAKV 36
QY 53 EYEYIVAF-NGYFTAKARNFISALKSSVD-NWRIIPRNPSDDYPSDFEVIQKEKQ 110
DB 37 EKDYIVGFKSGVTSVKKDIKES--GKVDKQFRIINAAKAKL----- 80
QY 111 KAGLLEDPHNTKRVTPQRKFRSLKYAESDPTVPCNETRWKQSSRPLRASLSLG 170
DB 81 KEALKEYKNDPDV-----AYVEED----- 99
QY 171 SGFWHATGRSSRLRAIPROVAQTLQADVLMQMGYTGANRVAVFDTLGSEKHPHFKN 230
DB 100 -----HVAHALAGTVYGI-PLIKADKVOAQEGKANRVAVLDTGIQASHPDINV 149
QY 231 VKERTNWTNERTLDGLGHGTFVAGVIASMREQQ---GFAPDAELHIFRVTNNQVSYS 287
DB 150 VGGASFVAGEAYNTDNGHGHVAGTVAALDNTTGVLGVPASVSLYAKVILNSSGSGTYS 209
QY 288 WFLDAFNAILKIDVNLNLSIGGPDFMD--HPFVQKVMELTANNVIMVSAIGNDGPL--Y 343
DB 210 GIVSGIEWATNGMDVNTNMSLGGPSGTAMKQAVDNAY---ARGVVVAAAAGNSGSGNT 266
QY 344 GTLNNPADQMDVIGVGIDFEDNTARFSSRGMTWELPGCYGRMKPDIVYGGVRSV 403
DB 267 NTGYPKYSVTAAGVADNSNRASFSSVGA---EL-----EVMAPGAGVYSTYP 314
QY 404 KGCRLSGTSVSPVAVAGAVTLLVSTVQKRLVNPASMKOALIASARRLPGVNMFPQGH 463
DB 315 TSYATLNGTSMASPHVAGAAALILS---KHPNLSASQVRNLSSTATYLG--SSFYVGK 369
QY 464 GKLDLLRAYQ 473
DB 370 GLINVEAAQ 379

RESULT 10

SUBT_BACAM
ID SUBT_BACAM STANDARD; PRT; 382 AA.
AC P00782;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Subtilisin BPN' precursor (EC 3.4.21.62) (Subtilisin Novo) (Alkaline
protease).
GN APR.
OS Bacillus amyloliquefaciens.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1390;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=ATCC 23844;
RX MEDLINE=85006739; PubMed=6090391;
RA Vasantha N., Thompson L.D., Rhodes C., Banner C., Nagle J.,
RA Filipula D.;
RT "Genes for alkaline protease and neutral protease from Bacillus
RT amyloliquefaciens contain a large open reading frame between the
RT regions coding for signal sequence and mature protein.";
RL J. Bacteriol. 159:811-819 (1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84069812; PubMed=6316278;
RA Wells J.A., Ferrari E., Henner D.J., Estell D.A., Chen E.Y.;
RT "Cloning, sequencing, and secretion of Bacillus amyloliquefaciens

subtilisin in Bacillus subtilis.";
RL Nucleic Acids Res. 11:7911-7925 (1983).
RN [3]
RP SEQUENCE OF 108-382.
RX MEDLINE=68086682; PubMed=6065094;
RA Markland F.S., Smith E.L.;
RT "Subtilisin BPN. VII. Isolation of cyanogen bromide peptides and the
RT complete amino acid sequence.";
RL J. Biol. Chem. 242:5198-5211 (1967).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
RX MEDLINE=72035041; PubMed=4399039;
RA Alden R.A., Wright C.S., Kraut J.;
RT "A hydrogen-bond network at the active site of subtilisin BPN'.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 257:119-124 (1970).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF COMPLEX WITH INHIBITOR.
RX MEDLINE=85033707; PubMed=6387152;
RA Hirono S., Akagawa H., Mitsui Y., Iitaka Y.;
RT "Crystal structure at 2.6-A resolution of the complex of subtilisin
RT BPN' with streptomyces subtilisin inhibitor.";
RL J. Mol. Biol. 178:389-413 (1984).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF MUTANT.
RX MEDLINE=90057412; PubMed=2684274;
RA Pantoliano M.W., Whitlow M., Wood J.F., Dodd S.W., Hardman K.D.,
RA Rollence M.L., Bryan P.N.;
RT "Large increases in general stability for subtilisin BPN' through
RT incremental changes in the free energy of unfolding.";
RL Biochemistry 28:7205-7213 (1989).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
RA Gallagher T., Oliver J., Bott R., Betzel C., Gilliland G.L.;
RT "Subtilisin BPN' at 1.6-A resolution: analysis for discrete disorder
RT and comparison of crystal forms.";
RL Acta Crystallogr. D 52:1125-1135 (1996).
RN [8]
RP ACTIVE SITE.
RX MEDLINE=69104413; PubMed=5249818;
RA Markland F.S., Shaw E., Smith E.L.;
RT "Identification of histidine 64 in the active site of subtilisin.";
RL Proc. Natl. Acad. Sci. U.S.A. 61:1440-1447 (1968).
CC -1- FUNCTION: SUBTILISIN IS AN EXTRACELLULAR ALKALINE SERINE PROTEASE,
CC IT CATALYZES THE HYDROLYSIS OF PROTEINS AND PEPTIDE AMIDES.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in PL. Hydrolyzes peptide amides.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
CC Alcalase by Novozymes.
CC -1- MISCELLANEOUS: SECRETION OF SUBTILISIN IS ASSOCIATED WITH ONSET OF
CC SPOULATION, AND MANY MUTATIONS WHICH BLOCK SPOULATION AT EARLY
CC STAGES AFFECT EXPRESSION LEVELS OF SUBTILISIN. HOWEVER, SUBTILISIN
CC IS NOT NECESSARY FOR NORMAL SPOULATION.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K02496; AAB05345.1; -
DR PIR; A00970; SUBSN.
DR PDB; 1S01; 15-OCT-90.
DR PDB; 1S02; 15-JAN-92.
DR PDB; 1SBH; 07-DEC-95.
DR PDB; 1SBI; 07-DEC-95.
DR PDB; 1SBN; 31-JAN-94.
DR PDB; 1SBT; 31-MAY-84.
DR PDB; 2SBT; 31-MAY-84.

FT CHAIN 150 1398 PYROLYSIN.
 FT ACT_SITE 179 365 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 365 365 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 441 441 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 152 222 N-LINKED (PROBABLE).
 FT CARBOHYD 222 222 N-LINKED (POTENTIAL).
 FT CARBOHYD 228 228 N-LINKED (POTENTIAL).
 FT CARBOHYD 240 240 N-LINKED (POTENTIAL).
 FT CARBOHYD 257 257 N-LINKED (POTENTIAL).
 FT CARBOHYD 262 262 N-LINKED (POTENTIAL).
 FT CARBOHYD 268 268 N-LINKED (POTENTIAL).
 FT CARBOHYD 327 327 N-LINKED (POTENTIAL).
 FT CARBOHYD 406 406 N-LINKED (POTENTIAL).
 FT CARBOHYD 651 651 N-LINKED (POTENTIAL).
 FT CARBOHYD 663 663 N-LINKED (POTENTIAL).
 FT CARBOHYD 739 739 N-LINKED (POTENTIAL).
 FT CARBOHYD 792 792 N-LINKED (POTENTIAL).
 FT CARBOHYD 893 893 N-LINKED (POTENTIAL).
 FT CARBOHYD 908 908 N-LINKED (POTENTIAL).
 FT CARBOHYD 917 917 N-LINKED (POTENTIAL).
 FT CARBOHYD 929 929 N-LINKED (POTENTIAL).
 FT CARBOHYD 1048 1048 N-LINKED (POTENTIAL).
 FT CARBOHYD 1056 1056 N-LINKED (POTENTIAL).
 FT CARBOHYD 1084 1084 N-LINKED (POTENTIAL).
 FT CARBOHYD 1117 1117 N-LINKED (POTENTIAL).
 FT CARBOHYD 1133 1133 N-LINKED (POTENTIAL).
 FT CARBOHYD 1140 1140 N-LINKED (POTENTIAL).
 FT CARBOHYD 1208 1208 N-LINKED (POTENTIAL).
 FT CARBOHYD 1233 1233 N-LINKED (POTENTIAL).
 FT CARBOHYD 1237 1237 N-LINKED (POTENTIAL).
 FT CARBOHYD 1332 1332 N-LINKED (POTENTIAL).
 FT CONFLICT 607 609 Y A -> PKP (IN REF. 1).
 FT CONFLICT 881 881 Y -> H (IN REF. 1).
 SQ SEQUENCE 1398 AA; 154474 MW; 3550873A27D56552 CRC64;

Query Match 5.1%; Score 285; DB 1; Length 1398;
 Best Local Similarity 20.5%; Pred. No. 4.2e-12;
 Matches 184; Conservative 115; Mismatches 316; Indels 284; Gaps 38;

QY 245 DGLGHTGTVAGTA-----SMRECGGAPDAE 271
 DB 361 DGHGCHTAVAGYDSNDAWDLMSYGEWFSRLYGRDYNVTTDTTQGVAPQA 420
 QY 272 LAIFPVFNNOVYTSWFLDAFNALIKKIDVLNLSIGGPFMDHPVD-----KWVE 324
 DB 421 IMAIRVLSRGSMWDIIEGMYTAATHGADVISMSLGS---NAPILDTDPESVAYDE 476
 QY 325 LITAN-NVIMVSAIGNDGLYGLTLNAPADQMDVIGV-----GIDFED----- 365
 DB 477 LTKYGVVFIAGNEGPGINISGPGVATKALTGAAVPIVGVYSQALGYPDIYGF 536
 QY 366 -----NIARFSRGMTHWELPGYGRMKPDIVYAGVRSYGVK--GGCRALSGTS 414
 DB 537 YFFPAYTNVRIAFSSRG-----PRIDGEIKPNVVFYGIYSSLPMTWGGADFMSTGS 590
 QY 415 VASPVAGATLLVTVQKREL-VNPASMKQALIASARLFG-----VNMFEQGHKL 466
 DB 591 MATPHVGVVALLISGAKAEGIYYPDIKKVLESCATWLEGDPYTGQYTDLDQGHVL 650
 QY 467 DLLRAYQILNSYKPOASLPSYIDLTECPVMPYCSQPIYGGMPTVVNVTILNMGVTV 526
 DB 651 NVTKSWELKA-----INGTTLPIVDHWADKS--YSDFAEVLGVVIRGLYARN 697
 QY 527 RIVDKPMOPVLPQNGDNIEVAFS-YSSVLW--PW-SGLAISIS---VTKKAASWEGIA 579
 DB 698 SIPDIVEN--HIKYGDIETFTFIATETWIKPFPVSVILENNTEFVLRYKYDVEGLE 755
 QY 580 QHVMIVASPAETESKGAQSTVTKPLKVKIIPTPRSKRVLMDQ-----YNNL 631
 DB 756 POLYIGRIIDPTTPIVIEDELLATVIPEKT-----TPENNYTLTVDINGEMVTHFF 811
 QY 632 RYPPG-----YFPR--DNLRMKNRDLWNGDHIHTN 660

DB 812 TYPEGVYLYAMITTYDYLGRPDGMFVFPYQDLYLPAAYSNPMPGNWELVWTG-----FN 867
 QY 661 FRDMYQHLRSRGYFVEVLGAPTC-----FDAQSY-----GTLIMV 696
 DB 868 FAPLYE-----SGFLVRIYGVETPSPVWVNTYLDITWTFSEIFNTIYAPINATLPI 923
 QY 697 D-----SEERYF-----PEELAKLRDVDN-----GLSLVIFSDWN----- 728
 DB 924 GLGTYNASVESVGDEFFKIGIEVPEGTAEKIRIGNPSPNSDLDLYLSDKGNLVALD 983
 QY 729 -----TSVMRKVKFYDENTRWMP-----DTGGAN----- 754
 DB 984 GNPTAEVEVVEYKPCVYIVVHGYSVYDENGPNPTTTTDLVQVTLNNGNKLDKDSI 1043
 QY 755 IPALNELLSW-----NMGFSGLYEGEFTLANIDMAYASGSIKAFEDGVVITQF-- 807
 DB 1044 ILGSNEVVYVANTITDRDHPYGVYSGIIEIRDEVYQDTNYSIAKIPITLVLDKADFV 1103
 QY 808 -----KDGLEVLKQETAVVENP-----ILGLYQIPABGGGRIVL-YGDSNCLDD 852
 DB 1104 GLTPAEGVLGEARNYTLIVKHALTLEPVNATVIGNTYLTIDENGTVTFTTAPTKLGS 1163
 QY 853 SHR--QKDCFWLLDALQ-----YTSYGVTPSPSLSHSGNRQPPSGAGSVTPERNE 901
 DB 1164 EITVIVKKNFNTLEKTFQITVSEPEITEEDINEPKLAMSS-----PEANATIVSVE 1217

RESULT 16
 ISP_BACCS STANDARD; PRT; 321 AA.
 AC P29140;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE Intracellular alkaline protease (EC 3.4.21.-).
 OC Bacillus clausii
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 EN NCBI_Taxid=79880;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
 RA Kato C., Nakano T., Yamamoto M., Horikoshi K.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; D10730; BAA01573.1;
 DR PIR; S27501; S27501.
 DR HSPP; P00782; I501.
 DR MEROPS; S08.030;
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE_ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Serine protease.
 FT ACT_SITE 49 49 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 230 250 CHARGE RELAY SYSTEM (BY SIMILARITY).
 SQ SEQUENCE 321 AA; 33747 MW; 621169DF104026 CRC64;

Query Match 5.1%; Score 284.5; DB 1; Length 321;
 Best Local Similarity 29.4%; Pred. No. 4.7e-13;
 Matches 95; Conservative 49; Mismatches 122; Indels 57; Gaps 14;


```
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
RA Kobayashi T., Ito S., Yamashita O.;
RT "Structure of a new alkaline serine protease (M-protease) from
RT Bacillus sp. KSM-K16."
RL Acta Crystallogr. D 51:199-206(1995).
RN [2]
RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
RX MEDLINE=95358832; PubMed=7632397;
RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
RA Koiike K., Kawai S., Ito S.;
RT "Purification and properties of an alkaline protease from
RT alkalophilic Bacillus sp. KSM-K16."
RL Appl. Microbiol. Biotechnol. 43:473-481(1995).
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR PDB: 1MPT; 22-JUN-94.
DR MEROPS: S08.010.
DR InterPro: IPR000209; Peptidase_S8.
DR Pfam: PF00082; Subtilisin.
DR PRINTS: PR00723; SUBTILASE_HIS.
DR PROSITE: PS00136; SUBTILASE_HIS.
DR PROSITE: PS00137; SUBTILASE_HIS.
DR PROSITE: PS00138; SUBTILASE_SER.
KW Hydrolyase; Serine protease; 3D-structure.
FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
SQ SEQUENCE 269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match 4.7%; Score 262.5; DB 1; Length 269;
Best Local Similarity 30.4%; Pred. No. 1.4e-11;
Matches 85; Conservative 46; Mismatches 124; Indels 25; Gaps 9;

QY 187 RAIPQVAQTLQADVLWQMGYTGANRVAVFDTGLSEKHPFKVKNKERTNWT-NERTLDD 245
Db 2 QQVPGGISR-VQAPAAHNRGLTSGVKVAVLDTGIS-THPDL-NIRGGASFVGPSTQD 58
QY 246 GLGCHTFVAGVIASRE---COGFAPDAELHIFRFTVNNQVSYTQWELDFNFAILKID 302
Db 59 GNGCHTHVAGTIAALNNSIGVLGVAPSAAELVAVKVLGASGSGSVSSIAQGLEWAGNGMH 118
QY 303 VLNLISGDPFMDHPFVDKVVWELTANNVIMVSAIGNDGLXTLNNPADOMDVIGVGID 362
Db 119 VANLSIGSPS-PSATLEQAVNSATSRGVLVVAASGNSG--GSISYPARYANAMAVGATD 175
QY 363 FEDNTARPSKGMTWELPGYGRMKPDIVTYGAGYRGSGVKGCGCRALSGTSSVSPVAVG 422
Db 176 QNNRASFSQVAGL-----DIVAPGVNVOSTYPCGSTYASLNGTSMATPHVAG 223
QY 423 ATVLIVSTVQKRELVPNPSKQALIASARLPGVNNFEQG 462
Db 224 VAAVLV---KKNFSPSWNVQIRNHLKNTATGLGNTNLYSG 260

RESULT 22
ID NISP_LACIA STANDARD; PRT: 682 AA.
AC Q07596;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nisin leader peptide processing serine protease nisp precursor
DE (EC 3.4.21.-).
GN NISP.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
```



```
QY 363 FEDNIARESSRGMTTWELPGYGRMKPDIVTYGAGVSGVGRGCRALSGTSVASPPWAG 422
DB 176 QNNRASEFOYQAGL-----DIVAPGVNQSTYPGSTVASLNGTSMATPHVAG 223
QY 423 AVTLVSVQKRELVPNPSMKQALIASARRLPQVGNFEOG 462
DB 224 AAALV---KQNPSSWSNVQIRNHLKNTATSLGSLNLYGSS 260

RESULT 24
ID SUBV_BACSU STANDARD; PRT; 806 AA.
AC P20141,1992 (rel. 24, Created)
DT 01-DEC-1992 (rel. 24, last sequence update)
DT 15-JUN-2002 (rel. 41, last annotation update)
DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
GN VPR OR IPA-45R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
RX MEDLINE=92041574; PubMed=1938892;
RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
RA Pero J.;
RT "Cloning and characterization of the gene for an additional
RT extracellular serine protease of Bacillus subtilis.";
RL J. Bacteriol. 173:6889-6895(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky E., Marcelino L., Moszer I.,
RA Prescan E., Santanu M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogatawa N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bortero M.G., Bessieres P., Holotin A., Borchert S.,
RA Boursier R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetalle D., Porwollik S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale I.,
RA Sato T., Scanlan E., Schleich S., Schroeter K., Scoffone F., Soldo B.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takanashi P., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosiato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Yoshikawa K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa K., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
```

```
subtilis.";
RT Nature 390:249-256(1997).
CC -!- FUNCTION: NOT REQUIRED FOR GROWTH OR SPORULATION.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PM: PROBABLY UNDERGOES C-TERMINAL PROCESSING OR PROTEOLYSIS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M76590; AAA22881.1; -
DR EMBL; X73124; CAA51601.1; -
DR EMBL; Z99123; CAB15835.1; -
DR PIR; A41341; A41341.
DR HSSP; P00782; 2SPT.
DR MEROPS; S08.00A; -
DR Subtilisin; BG10591; vpr.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 2.
DR Pfam; PF02225; PA; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 28
FT PROPEP 29 160
FT CHAIN 161 806
FT ACT_SITE 189 189
FT ACT_SITE 233 233
FT ACT_SITE 534 534
FT SEQUENCE 806 AA; 85608 MW; F984E3BF0B869DDDD CRC64;
SQ
Query Match 4.48; Score 249; DB 1; Length 806;
Best Local Similarity 20.2%; Pred. No. 7e-10; Indels 278; Gaps 22;
Matches 130; Conservative 69; Mismatches 166;
QY 25 LEKGFEPKAPCGCSHLTLKVEFSSTVVEYEVAFNGYFTAKARNSFISSALKSSEVDN 84
DB 62 LKESLAKAKGAGESQSKLKTART-----KAKNAI-KAVKNGKYNR 104
QY 85 WRIIPRNPSSDYFDEVIQIKKEKAGLLTLEHDPNKRKVTQPKRVSRLKYAE---S 141
DB 105 -----EYQVFGFSMK-----LPANEIPKLLAVKDVKAVYPNVYTKDNMK 146
QY 142 DPTVPCNTRWSQKWSRPLRRASLSLGSFGFWHATGRHSRRLLRAIPRQAQTLQADV 201
DB 147 KDKVTISEDVSPQMDSDAPY-----IGAND 172
QY 202 LQMCGYTCANVRVAVFDTGLSEKHPKKNKERTNWTNRTLD----- 244
DB 173 ANDLGYTGKFKVAIDTGYEYNHPLKK-----NFGYKGYDFVNDYDPKETPTGDP 227
QY 245 -DGLGCHTFVAGVITASNRCECFADPAELHIFRVFTNNQVSYTSWFLDAFNAILKKTVD 303
DB 228 GEATDGTGHTVAGTVAANGIKGVADPAELLAYRVLPGSGGTTENVIAQVRAVDGADV 287
QY 304 LMLSTG-----GPDPMDBPFVDKRWELTANNIMVISAINDGPLYGLTNNPADMDVIGV 359
DB 288 NMLSLGSLNPNPDWATSTALD---WAM-SEGVAVTSGNSGPNMGWVGSPGTSREAIISVQ 344.
QY 360 -----GI-----DFE----- 364
DB 345 ATQPLPNEYAVTFGSYSSAKVMGYNKEDDVKALNNKEVELVEAGIGEAQDFEGKDLTGKV 404
QY 365 ----- 364
```

Db 405 AVVKGSGIAFVADKADNAKAGAGMVMVYNNLSGEIANVPGMSVPTIKLSLEDGEKLVSA 464
QY 365 -----DNIAFFSRG--MTWELPGYGRMKPDIVTYGACV----- 398
Db 465 LKAGETKTKLTVSKALGQVADFSRGPVMDTW-----MIKPDISAPGVNIVSTIP 517
QY 399 -----RSGVGKGCRCALSGTSVASPVAGAVTLVSTVQKRELVPASMKQALIASAR 451
Db 518 THDPDHPYGVGSK-----QGTSMASPHIAGAVAVIKQAKPKWSV---EQIKAAIMTAV 568
QY 452 RLPGVN-----MFEQHGKLDLLRAYOILNSYRQASLSP-SY 488
Db 569 TLKSDSGEVYPHNAQAG-----SARIMNAIKADSLVSPGSY 605

RESULT 25

SUBE_BACSU STANDARD; PRT; 645 AA.

AC P16396;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Minor extracellular protease EPR precursor (EC 3.4.21.-).
GN EPR OR IPA-15R.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / DB204;
RX MEDLINE=90340301; PubMed=2116590;
RA Brueckner R., Shoseyov O., Doi R.H.;
RT "Multiple active forms of a novel serine protease from Bacillus
subtilis".
RL Mol. Gen. Genet. 221:486-490(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89053875; PubMed=3142851;
RA Sloma A., Ally A., Ally D., Pero J.;
RT "Gene encoding a minor extracellular protease in Bacillus subtilis.";
RL J. Bacteriol. 170:5557-5563(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S9.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: X53307; CAA37392.1;
EMBL: M22407; AAA2423.1;
EMBL: X73124; CAA51571.1;
EMBL: Z99123; CAB15866.1;
PIR: S11504; SUBSMP.
PIR: S39670; S39670.
HSP: O99405; IMPT.
MEROPS: S08.0PA; -.
Subtilist: BG10561; epr.
InterPro: IPR000209; Peptidase_S8.

DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; zymogen; Signal; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 103 POTENTIAL.
FT CHAIN 104 645 MINOR EXTRACELLULAR PROTEASE EPR.
FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 172 172 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 645 AA; 69695 MW; 401A4D5B60BE24A CRC64;

Query Match 4.38; Score 243; DB 1; Length 645;
Best Local Similarity 25.18; Pred. No. 1.3e-09;
Matches 104; Conservative 64; Mismatches 156; Indels 90; Gaps 19;

QY 99 SDFEVIQIKKQKAGLLTLED-----HPNIKRVTPQKVFSLKYAESDPTVPCN 148
Db 31 SEREVI-VVYKNKAGKETILDSADVQYKHLPAVAVTADQETVKELK---QDPDILYV 86
QY 149 ETRWSQKWQSRPLRRASLSLGSFHWATGRHSRRLLRAIPROVAOTLQADVLMQGYT 208
Db 87 ENNVSTAAADSTDFKVLK-----DGTDTSDNFEQWNLPEIQVQA-----WKAGLT 132
QY 209 GANVRVAVFDGTLSEKHPH-----FKNVKERTNNTNERTLDDGLGHGTFVAGVIASM 260
Db 133 GKNIKIAVIDSGIS---PHDDLSIAGYSVSYTSYKDDN-----GHGTHVAGIIGAK 183
QY 261 RE---COGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKIDVLNLSTG--GPDFMD 315
Db 184 HNGVGIDGIAPAEQIYAVKALDQNGSGDLSLQIDWSIANRMDIVNMSLGTTSDSKIL 243
QY 316 HPFVDKVMWELTANNVIMVSAIGNDG---PLYGTINNADQMDVIGVGIGIDFEDNIARPS 372
Db 244 HDVANKAYE---QGVLLVAASGNDGNGKP---VNYPAAYSSVVAVSATNEKNOLASEST 296
QY 373 RGMTTWELPGYGRMKPDIVTY---GAGVRGSGYKGCRCALSGTSVASPVVAGAVTLLVS 429
Db 297 TG-----DEVEFSAPGNTIISTYLNQYATGSGISQATPHRAAMFALL-- 339
QY 430 TVOKRELVPASMKQALIASARLPGVNMFEQHGKLDLLRAYOILNSYKPKQAS 483
Db 340 --KQRD--PAETNVQLREEMRK----NIVDLGTAGRDQQQGYGLI-QYKAQAT 383

Search completed: May 29, 2003, 13:38:54

Job time : 30 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 29, 2003, 13:33:32 ; Search time 50 seconds
(without alignments)
2022.668 Million cell updates/sec

Title: us-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIWLLLVLLCGKKH.....PRVKRQLMQVHPKTPSV 1052

Scoring table: BLOSUMP2

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 73.3

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5487	97.7	1052	T17093	intraluminal subti
2	1716	30.6	318	T43492	hypothetical prote
3	340	6.1	381	JQ1487	subtilisin (EC 3.4
4	338.5	6.0	381	1 SUBSS	subtilisin (EC 3.4
5	338.5	6.0	381	2 JH0778	subtilisin (EC 3.4
6	337.5	6.0	381	1 SUBSI	subtilisin (EC 3.4
7	329.5	5.9	379	1 SUBSCL	serine proteinase,
8	327.5	5.8	627	2 D75393	subtilisin (EC 3.4
9	322	5.7	382	2 I39780	subtilisin (EC 3.4
10	319.5	5.7	382	1 SUBSN	subtilisin (EC 3.4
11	312.5	5.6	272	2 A23624	subtilisin (EC 3.4
12	308	5.5	274	1 SUBSD	subtilisin (EC 3.4
13	304.5	5.4	323	2 I39867	microbial serine p
14	299	5.3	326	1 C41335	subtilisin (EC 3.4
15	295	5.3	275	2 JC1085	intracellular alka
16	294	5.2	321	2 JC5460	microbial serine p
17	293.5	5.2	322	2 G83922	subtilisin (EC 3.4
18	289	5.1	319	2 I39866	intracellular alka
19	285.5	5.1	374	2 I39781	microbial serine p
20	284.5	5.1	321	1 S27501	subtilisin (EC 3.4
21	283	5.0	380	2 A49778	alkaline proteins
22	281.5	5.0	1398	2 T28159	high-alkaline seri
23	276.5	4.9	384	2 C4802	pyrolysin (EC 3.4
24	270.5	4.8	328	2 G95392	alkaline proteinas
25	266	4.7	378	2 A33973	probable protease
26	260.5	4.6	682	2 S44131	high-alkaline seri
27	259	4.6	757	2 C84120	subtilisin-like pr
28	257	4.6	1448	2 A12007	subtilisin-type pr
29	253	4.5	397	2 JW0075	Subtilase family p
					cysteine-dependent

30	251.5	4.5	436	2	I39973	alkaline serine pr
31	250	4.5	442	2	A69587	intracellular alka
32	249.5	4.4	1345	2	T29090	surface layer-asso
33	249	4.4	806	2	A41341	microbial serine p
34	246	4.4	1331	2	A72647	probable surface l
35	243.5	4.3	401	2	A57690	aerolysin precurs
36	243	4.3	645	1	SUBSMP	serine proteinase
37	242	4.3	419	1	S25835	subtilisin (EC 3.4
38	241.5	4.3	412	2	T42024	probable serine pr
39	241.5	4.3	420	1	S23407	subtilisin (EC 3.4
40	241.5	4.3	792	2	H83736	subtilisin-type al
41	240	4.3	401	2	I39974	serine proteinase
42	239	4.3	444	2	B83891	intracellular alka
43	236	4.2	372	2	D83735	subtilisin-type al
44	235	4.2	488	2	A11930	proteinase [import
45	231.5	4.1	799	2	G83753	subtilisin-type pr

ALIGNMENTS

RESULT 1

T17093

Intraluminal subtilisin-like proteinase SLP, membrane-bound - Chinese hamster
C:Species: Cricetus griseus (Chinese hamster)
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T17093

R:Sakai, J.; Rawson, R.B.; Espenshade, P.J.; Cheng, D.; Seegmiller, A.C.; Goldstein,

Molecular Cell 2, 505-514, 1998

A:Title: Molecular identification of the sterol-regulated luminal protease that cleav

A:Reference number: Z18677; MOID:99026600; PMID:9809072

A:Accession: T17093

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-1052 <SA>

A:Cross-references: EMBL:AF078105; NID:G3892203; PID:G3892204; PIDN:AAC78321.1

A:Experimental source: strain 25-RA

C:Function:

A:Description: SLP is proposed to be a sterol-regulated protease that controls lipid

A>Note: SLP cleaves sterol regulatory element binding proteins (SREBPs) in the ER lum

Query Match 97.7%; Score 5487; DB 2; Length 1052;
Best Local Similarity 97.1%; Pred. No. 0;
Matches 1021; Conservative 16; Mismatches 15; Indels 0; Gaps 0;

QY	1	MKLVNIWLLLVLLCGKKHLDRLKKSFERAPCCSHLTLLKVFSSVVEYIVAF	60
DB	1	MKLVNIWLLLVLLCGKKHLDRLKKSFERAPCCSHLTLLKVFSSVVEYIVAF	60
QY	61	NGYFTAKARNSFTISSALKSEVDNWRITPRNPPSSDYPSPDFEVIQIKKQKAGLLTLEH	120
DB	61	NGYFTAKARNSFTISSALKSEVDNWRITPRNPPSSDYPSPDFEVIQIKKQKAGLLTLEH	120
QY	121	PNIKRVTPORKVFRSLKYAESDTPVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH	180
DB	121	PNIKRVTPORKVFRSLKYAESDTPVPCNETRWSQKWSRPLRRASLSLGSFHWATGRH	180
QY	181	SSRELLRATPROVAOTLQADVLWQMGYTGANVRVAVFDGLSEKHDFKVKERTWNTNE	240
DB	181	SSRELLRATPROVAOTLQADVLWQMGYTGANVRVAVFDGLSEKHDFKVKERTWNTNE	240
QY	241	RTLDGGLGHGTFVAGVIASMRCCQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK	300
DB	241	RTLDGGLGHGTFVAGVIASMRCCQGFAPDAELHIFRVFTNNQVSYTSWFLDAFNAILKK	300
QY	301	IDVLNLSIGGPDFMDHPFYDKYWELTANNVIMVSAIGNDGLYGLTNNPADQMDYGVGG	360
DB	301	IDVLNLSIGGPDFMDHPFYDKYWELTANNVIMVSAIGNDGLYGLTNNPADQMDYGVGG	360
QY	361	IDFENIARFSSRGMTTWELPGYGRMKPDIVYAGVRGSGVKGCGCRALSCTSVASPVV	420
DB	361	IDFENIARFSSRGMTTWELPGYGRMKPDIVYAGVRGSGVKGCGCRALSCTSVASPVV	420

QY 421 AGAVTLLVSTVQKRELVPASMKQALIASARRLPQGVNMFEGHKGKLDLRLAYOILNSYKP 480
DB 421 AGAVTLLVSTVQKRELVPASVKQALIASARRLPQGVNMFEGHKGKLDLRLAYOILNSYKP 480
QY 481 QASLSPSYIDTCEPYMPCYQPIYTGMPYVNTYVNTILNGMGVTRGVDPKPDWQPYLPQ 540
DB 481 QASLSPSYIDTCEPYMPCYQPIYTGMPYVNTYVNTILNGMGVTRGVDPKPDWQPYLPQ 540
QY 541 NGDNIEAFYSVSLWPNWSGYLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
DB 541 NGDNIEAFYSVSLWPNWSGYLAISIVTKKAASWEGIAQGHVMTIVASPAETESKNGAE 600
QY 601 QSTVKLPKIKVPIPTPPRSKRVLDQYHNLRYPPGYFPRDLNRMKNDPLDNGDHHTN 660
DB 601 HSTVKLPKIKVPIPTPPRSKRVLDQYHNLRYPPGYFPRDLNRMKNDPLDNGDHHTN 660
QY 661 FRDMYQHLRSMGYFVEVLGAPETCFDASOYGTLLMVDSEEFPEETAKLRDNDGLSL 720
DB 661 FRDMYQHLRSMGYFVEVLGAPETCFDASOYGTLLMVDSEEFPEETAKLRDNDGLSL 720
QY 721 VIFSDMYNTSVMBKVFYDENTRQWMPDGTGGANIPALNELLSSVWNGFSDGLYEGETL 780
DB 721 VIFSDMYNTSVMBKVFYDENTRQWMPDGTGGANIPALNELLSSVWNGFSDGLYEGETL 780
QY 781 ANHDMYASGCSIAKPPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
DB 781 ANHDMYASGCSIAKPPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGGGR 840
QY 841 IVLYGDSNCLDSDHROKDFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERM 900
DB 841 IVLYGDSNCLDSDHROKDFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERM 900
QY 901 EGNHLYRSKVLBAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHKLLSIDLDQV 960
DB 901 EGNHLYRSKVLBAHLGDPKPRLPACPRLSNAKPOPLNETAPSNLWKHKLLSIDLDQV 960
QY 961 VLPNFRSNRQVRLSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVYQ 1020
DB 961 VLPNFRSNRQVRLSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVYQ 1020
QY 1021 INKAKSRPKRRKPRVRPQMQVHPKPTSV 1052
DB 1021 ISKAKSRPKRRKPRVRPQMQVHPKPTSV 1052

RESULT 2
T43492
hypothetical protein DKEZp434A219.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C:Accession: T43492
R:Poushtka A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, December 1999
A:Reference number: 22316
A:Accession: T43492
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-318 <AAA>
A:Cross-references: EMBL:AL133583
A:Experimental source: adult testis; clone DKEZp434A219
C:Genetics:
A:Note: DKEZp434A219.1

Query Match 30.6%; Score 1716; DB 2; Length 318;
Best Local Similarity 100.0%; Pred. No. 5.2e-115;
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 735 VKFYDENTRQWMPDGTGGANIPALNELLSSVWNGFSDGLYEGETLANHDMYASGCSIA 794
DB 1 VKFYDENTRQWMPDGTGGANIPALNELLSSVWNGFSDGLYEGETLANHDMYASGCSIA 60
QY 795 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGGGRIVLYGDSNCLDSDH 854
DB 795 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGGGRIVLYGDSNCLDSDH 854

DB 61 KPEDGVVITQTFKDGLEVLKQETAVVENVPILGLYQIPAEGGGRIVLYGDSNCLDSDH 120
QY 855 ROKDCFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERMGNHLHRYSKVLEA 914
DB 121 ROKDCFWLLDALLQYTSYGVTTPSLSHSGNRQPPSPGAGSVTPERMGNHLHRYSKVLEA 180
QY 915 HLGDKPRLPACPRLSNAKPOPLNETAPSNLWKHKLLSIDLDQVLPNFRSNRQVRRP 974
DB 181 HLGDKPRLPACPRLSNAKPOPLNETAPSNLWKHKLLSIDLDQVLPNFRSNRQVRRP 240
QY 975 LSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVQINKAKSRPKRRKPR 1034
DB 241 LSPGSGAWDIPGGIMPGRYNOEGVQTIPIVFAFLGAMVLAFFVQINKAKSRPKRRKPR 300
QY 1035 VRRPOLMOOVHPKPTSV 1052
DB 301 VRRPOLMOOVHPKPTSV 318

RESULT 3
JQ1487
subtilisin (EC 3.4.21.62) J precursor - Bacillus stearothermophilus
C:Species: Bacillus stearothermophilus
C:Date: 17-Jul-1992 #sequence_revision 17-Jul-1992 #text_change 22-Jun-1999
C:Accession: JQ1487
R:Jang, J.S.; Kang, D.O.; Chun, M.J.; Byun, S.M.
Biochem. Biophys. Res. Commun. 184, 277-282, 1992
A:Title: Molecular cloning of a subtilisin J gene from Bacillus stearothermophilus
A:Reference number: JQ1487; MUID:92231938; PMID:1567435
A:Accession: JQ1487
A:Molecule type: DNA
A:Residues: 1-381 <JNA>
A:Cross-references: GB:M64743; NID:g142531; PIDN:AAA22247.1; PID:g142532
A:Experimental source: strain NCIMB10278
C:Genetics:
A:Gene: aprJ
A:Start codon: GTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-106/Domain: activation peptide #status predicted <ACT>
F:107-381/Product: subtilisin J #status predicted <MAT>
F:129-341/Domain: subtilisin homology <SBT>
F:138,170,327/Active site: Asp, His, Ser #status predicted

Query Match 6.1%; Score 340; DB 2; Length 381;
Best Local Similarity 25.1%; Pred. No. 2e-16;
Matches 123; Conservative 71; Mismatches 169; Indels 128; Gaps 18;

QY 1 MKLVNIWLLLVLLVLCGKHLGDRLEKKSFEKAPCCSHLTKVEFSFVY-----52
DB 1 MKLVNIWLLLVLLVLCGKHLGDRLEKKSFEKAPCCSHLTKVEFSFVY-----52

QY 53 EYEYIVAFNGYFTAKARNSFISALKSEVDNWRILIRNPNPSSDYPSDFEVIQKEKQKA 112
DB 53 EYEYIVAFNGYFTAKARNSFISALKSEVDNWRILIRNPNPSSDYPSDFEVIQKEKQKA 112

QY 36 EKKYIVGF-----KQTSAMSSAKKDDVI-----SEK 63
DB 36 EKKYIVGF-----KQTSAMSSAKKDDVI-----SEK 63

QY 113 GLLTLE-DHPNIRKRVTPQRKVPFRSLKYAESDPTVPCNETRWSKQWSSRPLRSLSGS 171
DB 113 GLLTLE-DHPNIRKRVTPQRKVPFRSLKYAESDPTVPCNETRWSKQWSSRPLRSLSGS 171

QY 64 GKVKQKRYVNAATAATLDEKAVKELK-----KDPSSVA-----95
DB 64 GKVKQKRYVNAATAATLDEKAVKELK-----KDPSSVA-----95

QY 172 GFHWATGRHSRRLLRAIPRQVAQTLQADVLWQMGYTGANVRVAVDTGLSEKHPFKNV 231
DB 172 GFHWATGRHSRRLLRAIPRQVAQTLQADVLWQMGYTGANVRVAVDTGLSEKHPFKNV 231

QY 96 ---VVEEDHIAHEYAQSVPYGISQ-IPALHSGQYTGSNVAVVIDSIDSSHPLD-NV 150
DB 96 ---VVEEDHIAHEYAQSVPYGISQ-IPALHSGQYTGSNVAVVIDSIDSSHPLD-NV 150

QY 232 KERTNWTNERT--LDLGLCHGTFAVGVIASHRE---COGFAPDAELHIFRFTNNQVSYT 286
DB 232 KERTNWTNERT--LDLGLCHGTFAVGVIASHRE---COGFAPDAELHIFRFTNNQVSYT 286

QY 151 RCGASFVPSPTNPDQSSSGHGTAVAGTAAALNNSIGVLGSPSASLYAVKVLDTSGSQY 210
DB 151 RCGASFVPSPTNPDQSSSGHGTAVAGTAAALNNSIGVLGSPSASLYAVKVLDTSGSQY 210

QY 287 SHWLDAPNVALKIDVNLISIGGDFMD--HPVDKVELTANNVIMVSAIGNDGP--L 342
DB 287 SHWLDAPNVALKIDVNLISIGGDFMD--HPVDKVELTANNVIMVSAIGNDGP--L 342

QY 211 SWLINGEIAISNMNDYNNKSLGGSPGSGTALKTVDKA---VSSGIVVAAGNAGSGSGS 267
DB 211 SWLINGEIAISNMNDYNNKSLGGSPGSGTALKTVDKA---VSSGIVVAAGNAGSGSGS 267

343	QY	YCTLANNADQMDVTCVCGIDEEDINAFSSRGMTTWELPGCYGRMPDVTIVTGAVGRGSS	402
268	Db	STVGYPKPYSTINAVGAVSSNRQAFSSAGS-----EL-----DYNAPGVSIQSTL	315
403	QY	VYGGCRALSGTSSPVPVAGAVTLLVSTVQKRELVPNPSMKQALIASARLLPGVNNFEOG	462
316	Db	PGGTGYGNGTSMATPHVAGAAAILLS---KHPTWTNAQVRDRLESTATYLG--NSFYFYG	370
463	QY	HGKLDLLRAYQ	473
371	Db	KGLINVAQAAQ	381

RESULT 4

SUBS

subtilisin (EC 3.4.21.62) amylosacchariticus precursor - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 24-Apr-1984 #sequence_revision 24-Feb-1995 #text_change 16-Jun-2000
C:Accession: A11448; A00971; S68013
R:Yoshimoto, T.; Oyama, H.; Honda, T.; Tone, H.; Takeshita, T.; Kamiyama, T.; Tsuru, D.
J. Biochem. 103, 1060-1065, 1988
A:Title: Cloning and expression of subtilisin amylosacchariticus gene.
A:Reference number: A11448; MWID:89008194; PMID:3139650
A:Accession: A11448
A:Molecule type: DNA
A:Residues: 1-381 <YOS>
A:Cross-references: GB:D00264; NID:9216328; PIDN:BAA00186.1; PID:9912425
A:Experimental source: var. amylosacchariticus
R:Kurihara, M.; Markland, F.S.; Smith, E.L.
J. Biol. Chem. 247, 5619-5631, 1972
A:Title: Subtilisin amylosacchariticus. III. Isolation and sequence of the chymotryptic
A:Reference number: A00971; MWID:72266688; PMID:5055784
A:Accession: A00971
A:Molecule type: protein
A:Residues: 107-112-114; 148-152; 155-157; 164-170; 173-174; 178-181; 200-205; 210-212; 219-225;
A:Experimental source: var. amylosacchariticus
R:Kamal, M.; Hooeog, J.O.; Kaiser, R.; Shafigat, J.; Razzaqi, T.; Zaidi, Z.H.; Joernvall, I.
FEBS Lett. 374, 363-366, 1995
A:Title: Isolation, characterization and structure of subtilisin from a thermostable Bac
A:Reference number: S68012; MWID:96069945; PMID:7589571
A:Accession: S68013
A>Status: preliminary
A:Molecule type: protein
A:Residues: 107-235; 'T', 237-245; 293-381 <KAM>
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many
not necessary for normal sporulation.
C:Genetics:
A:Start codon: GFG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:107-381/Product: subtilisin #status predicted <MAT>
F:129-341/Domain: subtilisin homology <SBT>
F:138-170,327/Active site: Asp, His, Ser #status predicted

Query Match	6.0%;	Score	338.5;	DB	1;	Length	.391;
Best Local Similarity	26.8%;	Pred. No.	2.6e-16;				
Matches	117;	Conservative	66;	Mismatches	156;	Indels	97;
Gaps	17;						
QY	48	SSTVVEYEVAFNGYFTAKARNSTFTSSALKSSEVDNWRLIPRNPSSDYPDFEVQIK	107				
				:			
Db	33	SST--EKKYVGF-----KQMSAMSSAKKQDVI-----	59				
QY	108	EKQKAGLLYLE-DHPNKRVTPOKYVRSUKYAESDTPVPCNTRWSQKWSSRPLRRAS	166				
				:			
Db	60	-SEKGGVKQKFYVNAATAATLDEKAVKELK----KDPSPA-----	95				
QY	167	LSLGSFWHATGHRSSRRLLRAIPROVAOTLQADVLWQMGYTCGANRVAVFDGLSEKHP	226				
				:			
Db	96	-----VVEEDHIAHEVAQSPYQISQ--IKAPALHSQGYTGSNNKVAVIDSGIDSSH	146				
QY	227	HFKNVKERTNWTNERT--LDDGLGHGTFVAGYIASMRE---CGFAPDAELHFRVFTNN	281				
				:			
Db	147	DI--NVRGGASFPVSETPNPYODSSHGCTHVAAGTIAALNNSIGVLGVSPASLYAVKVLDS	205				

Qy	282	QVSYTSWFLDAFNFAILKLDVNLSTGGPDM--HFFVKVWELTANNVIMWSAIGND	339
Db	206	GSQYISWIIINGIEWAINNMDVINMSLGGPSGSLTKTVDKA---VSSGIVVAAAAGNE	262
Qy	340	GP--LYLTNNPADQMDVIGVGGTDFEDNIARESSRGTTWELPGYGRMKRPDIYTYGAG	397
Db	263	GSQGSSTVGYPKYPTIAVGAVNSSQRASFSSAGS---EL-----DVNAPGVS	310
Qy	398	VRSGVGGCGGRALSGTSVASPVVAGVTLTVSTVQKRELVPASMKQALIASARLPGVN	457
Db	311	IQSTLPGGTYGAYNGTSMATPHVAGAAALILS---KHPWTNAQVDRDLRESTATYLG--N	365
Qy	458	MFEQGHGKDLLRAYQ	473
Db	366	SFYYGKGLINVQAAQ	381

RESULT 5

JH0778

subtilisin (EC 3.4.21.62) NAT precursor -Bacillus subtilis (strain natto NC2-1)
N:Alternate names: natto proteinase; nattokinase; subtilisin BSP
C:Species: Bacillus subtilis
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: JH0778; JS0601; JS0517; JC2036
R:Nakamura, T.; Yamagata, Y.; Ichishima E.
Biosci. Biotechnol. Biochem. 56, 1869-1871, 1992
A:Title: Nucleotide sequence of the subtilisin NAT gene, aprN, of Bacillus subtilis {
A:Reference number: JH0778; MUID:93113095; PMID:1369081
A:Accession: JH0778
A:Molecule type: DNA
A:Residues: 1-381 <NAK>
A:Cross-references: GB:D25319; NID:g435439; PIDN:BA04989.1; PID:g435440
R:Sumi, H.
Kagaku To Seibutsu 29, 119-123, 1991
A:Title: Natto kinase and fibrinolysis.
A:Reference number: JS0601
A:Accession: JS0601
A:Molecule type: protein
A:Residues: 107-381 <SUM>
R:Sumi, H.; Nakajima, N.
Nippon Nogelkagaku Kaishi 65, 1125-1127, 1991
A:Title: Studies on fibrinolysis enzymes in fermentation food.

```
Query Match          6.08; Score 338.5; DB 2; Length 381;
Best Local Similarity 27.1%; Pred. No. 2.6e-16;
Matches 118; Conservative 65; Mismatches 156; Indels 97; Gaps 17

QY 48 SSTVVEYEYVAENGFTAKARNFFISSALKSESDNWRRIIPRNPSDDYDFEVIQIK 107
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   33 SSI--EKKYIVGF-----KOTMSAMGSAKKKDVI----- 59
```

Qy		108	EKQKAGLITLE-DHPNIMIKRVTPQPKVFRSLKYAESDPTVPCNTNRWSOKWOSSRPLRRAS	166
Db		60	-SEGGKVQQRKYNNAAATLDEKAVKEK-----KDPFA-	95
Qy		167	LSIGSGFWHATGHRHSRRLLRAIPROVATGLQADVLMOMGYTCANVRVAVPDTGLSEKH	226
Db		96	-----YVEEDHTAHEYAOSVPYGISO- IKAPALHSQGTYGSNVKAVIDSGIDSSH	146
Qy		227	HFKNVKERTNWTNRT--LDGLGHGTGFVAGVIASMRE--COGFAPDAELHIFRVFTNN	281
Db		147	DV-NVRGASFPVSETNPYODSGSHGTWAGTIAALNNISGVLGVAAPSASYAKVLVDST	205
Qy		282	QVSYTSNFLDAFNATLKIDVLNLSTGGSPFMD--HPFDVKVWEUETANNIWMYSATGD	339
Db		206	GSGQYSWTLINGIEWAISNNMOVINNLGGTGTTALKTVVDKA--VSJGVVAAAAAGNE	262
Qy		340	GP-LYGTFLNPNADQMIVGVGGIDFEDNIARFSRGMTTWELPGCYGRMKPDVTIVYAG	397
Db		263	GSSGSTSVGYPAKYPSTIAYGAVNSNQARSFSSVGSS--EL-----DYMAPGVS	310
Qy		398	VFGSGVGGCRALSQTGSVASPVWAGAVLTLLVSTVQKRELYNPASKMOALLIASARRLPGVN	457
Db		311	IQLSTLPGGTYGANGTSMATPHVGAAGAAILSS---KHPTWTNAOVDRDLRESTATYLQ--N	365
Qy		458	MFEQHGKGLDLLRAYQ	473
Db		366	SFYVYKGKGLINVOAAAQ	381

RESULT 6

SUBS

subtilisin (EC 3.4.21.62) E precursor - *Bacillus subtilis*
N:Alternate names: alkaline proteinase; bacillopeptidase E; extracellular alkaline serin
C:Species: *Bacillus subtilis*
C:Date: 04-Dec-1986 #sequence revision 04-Dec-1986 #text change 16-Jun-2000
C:Accession: A00972; A26116; 139970; 139778; 139779; S68012; H69586
R:Stahl, M.B.; Ferrari, E.
J. Bacteriol. 158, 411-418, 1984
A:Ritle: Replacement of the *Bacillus subtilis* subtilisin structural gene with an in vitro
A:Reference number: A00972; MUID:84212198; PMID:6427178
A:Accession: A00972
A:Molecule type: DNA
A:Residues: 1-381 <STA>
A:Cross-references: GB:K01988; NID:g143519; PIDN:AAA22742.1; PID:g143520
R:Experimental source: strain 1168
R:Wong, S.L.; Price, C.W.; Goldfarb, D.S.; Doi, R.H.
Proc. Natl. Acad. Sci. U.S.A. 81, 1184-1188, 1984
A:Ritle: The subtilisin E gene of *Bacillus subtilis* is transcribed from a sigma37 promoter
A:Reference number: A26116; MUID:84144862; PMID:6322190
A:Accession: A26116
A:Molecule type: DNA
A:Residues: 1-155 <MON>
A:Cross-references: GB:K01443; NID:g143665; PIDN:AAA22814.1; PID:g143666
R:Ikenura, H.; Takagi, H.; Inouye, M.
J. Biol. Chem. 262, 7859-7864, 1987
A:Title: Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*
A:Reference number: 139969; MUID:87222417; PMID:3108260
A:Accession: 139970
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-156 <LKE>
A:Cross-references: GB:M16639; NID:g143521; PIDN:AAA22744.1; PID:g143523
R:Henner, D.J.; Ferrari, E.; Perego, M.; Hoch, J.A.
J. Bacteriol. 170, 296-300, 1988
A:Ritle: Location of the targets of the hpr-97, sacU32(Hy), and sacQ36(Hy) mutations in *Escherichia coli*
A:Reference number: 139778; MUID:88086885; PMID:2447063
A:Accession: 139778
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-8 <CHN>
A:Cross-references: GB:M19125; NID:g142527; PIDN:AAA22245.1; PID:g142528
R:Park, S.
J. Bacteriol. 171, 2657-2665, 1989

A>Title: Bacillus subtilis subtilisin gene (apre) is expressed from a sigma-A (sigmaA)
A:Reference number: I39779; MUID:89213955; PMID:2496113
A:Accession: I39779
A:Molecule type: DNA
A:Residues: 1-13 <PAR>
A:Cross-references: GB:M31060; NID:g142529; PIDN:AAA22246.1; PID:g142530
A:Experimental source: strain W168, substrain PY79
R:Ramal, M.; Hoeoeg, J.O.; Kaiser, R.; Shafqat, J.B.; 363-366, 1995
FERS Lett. 374,
A>Title: Isolation, characterization and structure of subtilisin from a thermostable
A:Reference number: S68012; MUID:96069945; PMID:7589571
A:Accession: S68012
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 113-323 <KAM>
R:Runst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Avevedo, V.; Ber
C.: Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.: Ehrlich, S.D.; Emmeron, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Cal
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holtsappell, S.; Hosono, S.; Hulio, S.;
Koester, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
Vier, M.; Ogawa, K.; Ogihara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete
Rleguer, M.; Rivolta, C.; Rocha, E.; Roche, M.; Sadate, Y.; Sato, T.; Scanl
A:Authors: Schleicher, S.; Schroter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstora, P.; Tognoni, A.; Tosato, V.; Uchiya
T.: Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumanst, E.; Yoshikawa, H.; Danchin, A.
A>Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili
A:Reference number: A65580; MUID:98044033; PMID:9384377
A:Accession: H69586
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-26, 'V', 28-381 <KUN>
A:Cross-references: GB:Z99109; GB:AL009126; NID:g2633260; PIDN:CAB12870.1; PID:g26333
A:Experimental source: strain 168
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
not necessary for normal sporulation.
C:Genetics:
A:Gene: apre
A:Map position: 690-771
A:Start codon: GTG
C:Function:
A:Description: catalyzes the hydrolysis of peptide bonds
A:Pathway: protein digestion
A>Note: This enzyme has broad specificity and will hydrolyze peptide amides; it prefer
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; protein digestion; serine proteinase
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-106/Domain: activation peptide #status predicted <APT>
F:107-381/Product: subtilisin E #status predicted <MPT>
F:129-341/Domain: subtilisin homology <SBT>
F:138, 170, 327/Active site: Asp, His, Ser #status predicted

K; Park, S.
J. Bacteriol. 171, 2657-2665, 1989

Db 147 DL-NVRGASFPVSETPNYQDSSGHGTHVAGTAAALNNSIGVLGSPSASLYAKVKLDST 205
QY 282 QVSVTSWFLDAENKAILKIDVLNLSIGGPDFMD--HPFVDKRWELTANNVIMVSAIGND 339
Db 206 GSGQSYWINGIEWATNSNMVDVIMSLGGPTGTALKTVVDKA---VSSGIVVAAAAGNE 262
QY 340 GP--LYGTLLNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAG 397
Db 263 GSSGSGTSTVGYPKYPSTIAGVAVNSNQASFSAGS---EL-----DVMAPGVS 310
QY 398 VSGGVKGGCRALSGTSVSPVAVAGVTLVSTVQRELNVNPSMKQALIASARRLPQVN 457
Db 311 IQSTLPGGTGAVNGTSMATPHVAGAAALILS---KHPTWNAQVRDRLESTATYLG--N 365
QY 458 MFEQGHGKDLLRAYQ 473
Db 366 SFYIGKGLINVQAAQ 381

RESULT 7

SUBSCL
subtilisin (EC 3.4.21.62) Carlsberg precursor - Bacillus licheniformis
C:Species: Bacillus licheniformis
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 21-Jul-2000
C:Accession: A24111; A00968
R:Jacobs, M.; Eliasson, M.; Uhlen, M.; Flock, J.I.
Nucleic Acids Res. 13, 8913-8926, 1985
A:Title: Cloning, sequencing and expression of subtilisin Carlsberg from Bacillus licheniformis
A:Reference number: A24111; MUID:86093688; PMID:3001653
A:Accession: A24111
A:Molecule type: DNA
A:Residues: 1-379 <JAC>
A:Cross-references: GB:X03341; NID:9487721; PIDN:CAB56500.1; PID:g5921206
A:Experimental source: strain NCIB6816
R:Smith, E.L.; DeLange, R.J.; Evans, W.H.; Landon, M.; Markland, F.S.
J. Biol. Chem. 243, 2184-2191, 1968
A:Title: Subtilisin Carlsberg. V. The complete sequence; comparison with subtilisin BPN'
A:Reference number: A00968; MUID:68234702; PMID:4967581
A:Accession: A00968
A:Molecule type: protein
A:Residues: 106-206, 'S', '234-261, 'N', '263-264, 'S', '266-315, 'N', '317-379 <SMT>
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and many not necessary for normal sporulation.
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-105/Domain: propeptide #status predicted <APT>
F:106-379/Product: subtilisin Carlsberg #status experimental <MPT>
F:128-339/Domain: subtilisin homology <SBT>
F:137,168,325/Active site: Asp, His, Ser #status predicted

Query Match 5.9%; Score 329.5; DB 1; Length 379;
Best Local Similarity 25.3%; Pred. No. 1.1e-15;
Matches 124; Conservative 66; Mismatches 171; Indels 129; Gaps 17;
QY 1 MKLVNIWLLLVLLVCKHGLDRLEKKEKAPCPGCSHLTLKVFSSTV-----V 52
Db 2 MRKKSFWLGMATFM-----LVFTMAFSDSASAAQAKNV 36
QY 53 EYEYIYAP-NGFTAKARNSFTISSALKSEVD-NWRIRPNNDSSYPSPDVEVIQKEQ 110
Db 37 EKDIYVGFSGKVTASVKKDIKES--GGKVDKQRIINAARAKLD----- 80
QY 111 KAGLLTLEDPNPKRVTPQKRVSRSLKYAESDTPVPCNETRWSQKWSRPLRASLSLG 170
Db 81 KEALVEKNDPDV-----AYVEED----- 99
QY 171 SGFWHATGRSRRLLRAIPROVAOTLQADVLWQMGYTGANVRVAVFDTGLSEKHPHEKN 230
Db 100 -----HVAHALAQTPYGI-PLIKADKVAQGFKANVKAVIDLGIOASHPDLVN 149
QY 231 VKERTNWNERTLDDGLGHTGVAGVIASMRQCG---GFAPDAELHIFRVFTNNQVSYTS 287

Db 150 VVGASFVAGEAYNTDGNHGHVAGTVAALDNTTGLVGVAPSVSLYAKVNLSSGSGTYS 209
QY 288 WFLDAFNAYLKKIDVLNLSIGGPDFMD--HPFVDKRWELTANNVIMVSAIGNDGPL--Y 343
Db 210 GIVSGIEWATNGMDVINMSLGGPSGTAMQAVDNAY---ARGVVVAAAAGNSGSGNT 266
QY 344 GTLNNPADQMDVIGVGIDFEDNIARFSSRGMTTWELPGGYGRMKPDIVTYGAGVRCGSV 403
Db 267 NTIGYPAKYDSVIAVGAVDNSNRASFSSVGA---EL-----EVMAPGAGVYSTYP 314
QY 404 KGCGRALSGTSVSPVAVAGVTLVSTVQRELNVNPSMKQALIASARRLPQVN 463
Db 315 TSTYATLNGTSMATPHVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSFYIGK 369
QY 464 GKDLLRAYQ 473
Db 370 GLINVEAAQ 379

RESULT 8

D75393
serine proteinase, subtilase family - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: D75393
R:White, O.; Eisen, J.A.; Heideberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: D75393
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <WHI>
A:Cross-references: GB:AE001990; GB:AE00513; NID:96459214; PIDN:AAF11026.1; PID:g645
A:Experimental source: strain R1
C:Genetics:
A:Gene: DRI459
A:Map position: 1

Query Match 5.8%; Score 327.5; DB 2; Length 627;
Best Local Similarity 27.5%; Pred. No. 3.5e-15;
Matches 119; Conservative 62; Mismatches 181; Indels 71; Gaps 16;
QY 57 IVAFNGVFTAKARNSF--ISSALKSEVDNWRIRPNNDSSYPSPDVEVIQKEKAGL 114
Db 47 IVRFNANTAQGRALFKNLRGQLNSQ-----IAKLGPSAGF-----LQAVNSQKATQ 94
QY 115 LTLEDHPNTRKRVTPQKRVSRSLKYAESDTPVPCNETRWSQKWSRPLRASLSLGSGFW 174
Db 95 LWLDDQSVLLPMTVQARAVSOLPFVAA-----VFENFKVQIPRVVAMSNASAPAG-- 144
QY 175 HATGRHSRRLLRAIPROVAOTLQADVLWQMGYTGANVRVAVFDTGLSEKHP----- 226
Db 145 --TPTH-----LQIGAPAAWAAAGFGKGNIRIGHLDSDIDPSHPELAGKVA 189
QY 227 --HFNKVKERTNWTNERTLDDGLGHTGVAGVIASMRQCGFAPDAELHIFRVFTNNQVS 284
Db 190 FQEFNGEGDVRSSQPHDITD-----HGHTAGLLVGSK--YGVAPGARVISALVLPNNEG 243
QY 285 YT-----SWFLDAFNAYLKK-IDVLNLSIGGPDFMDHFPVDKRWELTANNVIMVSAIG 337
Db 244 FAQVIAGMQVLDPDNDADDDGADVNNMSLGLPGTWNE--FIYVNNMLKAGVVPVFAIG 302
QY 338 NDGPXYGLTNLPADQMDVIGVGIDFEDNIARFSSRGMTTW--ELPGGYGRMKPDIVTYG 395
Db 303 NFGPAAGSTGSPCNLPAIGVGAVDNSNQVAFSSSRGPPVWQGEISSVF--TRPDIAAPG 360
QY 396 AGVRGSGVGGCRALSGTSVSPVAVAGVTLVSTVQRELNVNPSMKQALIASARRLP 455
Db 361 VNITSVRNGGYQAMSGSGSQASPTAGAVALLSA---KPGASVDAIKNALFTSASNA 417

QY 456 VNMFEQGHGKLDL 468
DB 418 KNN-NVGFQGISI 429

RESULT 9
139780
subtilisin (BC 3.4.21.62) Sendai precursor - Bacillus sp.
C:Species: Bacillus sp.
C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 22-Jun-1999
C:Accession: I39780
R:Yamagata, Y.; Ishihara, K.; Ichishima, E.
Enzyme Microb. Technol. 17, 653-663, 1995
A:Title: Subtilisin Sendai from alkalophilic Bacillus sp.: molecular and enzymatic properties
A:Reference number: I39780; MUID:95329264; PMID:7605625
A:Accession: I39780
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-382 <RES>
A:Cross-references: GB:D29688; NID:9995963; PIDN:BA06157.1; PID:9995964
C:Genetics:
A:Start codon: TTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:136-342/Domain: subtilisin homology <SBT>

Query Match 5.7%; Score 322; DB 2; Length 382;
Best Local Similarity 27.4%; Pred. No. 3.9e-15;
Matches 124; Conservative 59; Mismatches 172; Indels 98; Gaps 19;

QY 27 KKSFKAPCCGSHLPLKVFSSVVEY-----YIVAFNGYFTAKARNSFTSSALK-SS 80
DB 2 KLEFTKVVASA--LLLSISLTATSVSAEQQKQYLVIGFENQLQV---TEFVSSDKGQS 56

QY 81 EVDNRIIPNPPSDYPSDFEVIQKEKAGLTLDHPNKRIPQPKVFSLKVAE 140
DB 57 EMSLFAEVNDESEMELLYEFEDIPVVSVE---LSPEDKALEK-DP-----SITVIE 105

QY 141 SD--PTVPCNETRWQKQSSRLRRASLSIGSGFWATGRHSSRRLLRAIPROVAQTLQ 198
DB 106 EDIEVTITNQVTPW-----GTR-----VQ 125

QY 199 ADVLWQMGYGANVRVAVFDGLSEKHPKVKERTWNT-NERTLDGLGHGTFVAGVI 257
DB 126 APTATRGYTGTVRVAVLDGIGS-THPDL-NIRGVSVFVCEPSYQDNGRGTHTVAGTI 183

QY 258 ASMRE---CGFPADAEHLHFRVFTNNQVSYTWFDAFNATIAILKIDVNLNLSIGGPDFM 314
DB 184 AALNNSGVVGVAPNAELFAYKVLGANGSGSVSSTAQGLQWTAONNLIHANLSLGS---- 239

QY 315 DHPFDVKWELTAN-----NVIMVSAIGNDGLYGLTLNPNADQMDVIGVGGIDFEDNAR 369
DB 240 --PVGSQTLLEYLNQATNAGVLVVAATNGNGS--GTVSYPARYANALAVGATDQNNRAS 295

QY 370 FSSRGMTTWELPGCYGRMKPDIVTYGAGVSGVKGCGCRALSGTSVASPVVAGAVTLVLS 429
DB 296 FSOXGTGL-----NIVAPGVGIQSTYPGNRYASLSGTSNATPHVGAALV-- 341

QY 430 TVQKRELVPASMKQALIASARRLPGVNMFEQ 462
DB 342 -KQKNPNSWNTQIRQHLTSTATSLGNSNQPGSG 373

RESULT 10
SUBSN
subtilisin (BC 3.4.21.62) BPN' precursor - Bacillus amyloliquefaciens
N:Alternate names: subtilisin Novo
C:Species: Bacillus amyloliquefaciens
C>Date: 24-Apr-1984 #sequence_revision 28-Aug-1985 #text_change 21-Jul-2000
C:Accession: B25415; A93495; T44584; A92033; A00970
R:Vasantha, N.; Thompson, L.D.; Rhodes, C.; Banner, C.; Nagle, J.; Filpula, D.
J. Bacteriol. 159, 811-819, 1984

A:Title: Genes for alkaline protease and neutral protease from Bacillus amyloliquefac
A:Reference number: A25415; MUID:85006739; PMID:6090391
A:Accession: B25415
A:Molecule type: DNA
A:Residues: 1-382 <VAS>
A:Cross-references: GB:K02496; NID:9142525; PIDN:AAB05345.1; PID:9142526
A:Experimental source: ATCC 23844
R:Wells, J.A.; Ferrari, E.; Henner, D.J.; Estell, D.A.; Chen, E.Y.
Nucleic Acids Res. 11, 7911-7925, 1983
A:Title: Cloning, sequencing, and secretion of Bacillus amyloliquefaciens subtilisin
A:Reference number: A93495; MUID:84069812; PMID:6316278
A:Accession: A93495
A:Molecule type: DNA
A:Residues: 1-382 <WEL>
A:Accession: T44584
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-382 <WEL>
A:Cross-references: EMBL:X00165; NID:939337; PIDN:CAA24990.1; PID:9773560
R:Markland, F.S.; Smith, E.H.
J. Biol. Chem. 242, 5198-5211, 1967
A:Title: Subtilisin BPN'. VII. Isolation of cyanogen bromide peptides and the complete
A:Reference number: A92033; MUID:68086882; PMID:6065094
A:Accession: A92033
A:Molecule type: protein
A:Residues: 108-162, 'PN', 165-167, 'D', 169-194, 'SA', 197-204, 'DA', 207-264, 'ST', 267-357, '
R:Kraut, J.
In The Enzymes, 3rd ed., vol.3, Boyer, P.D., ed., pp.547-560, Academic Press, New York
A:Title: Subtilisin: X-ray structure.
A:Reference number: A94443
C:Contents: annotation; X-ray crystallography, 2.5 angstroms; active site
C:Comment: Secretion of subtilisin is associated with the onset of sporulation, and m
not necessary for normal sporulation.
C:Genetics:
A:Start codon: GTG
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-32/Domain: signal sequence #status predicted <SIG>
F:33-107/Domain: activation peptide #status predicted <APT>
F:108-382/Product: subtilisin BPN' #status experimental <MPT>
F:130-342/Domain: subtilisin homology <SBT>
F:139,171,328/Active site: Asp, His, Ser #status experimental

Query Match 5.7%; Score 319.5; DB 1; Length 382;
Best Local Similarity 31.7%; Pred. No. 5.9e-15;
Matches 98; Conservative 50; Mismatches 118; Indels 43; Gaps 13;

QY 180' HSSRLLRAIPROVAQTLQADVLWQMGYGANVRVAVFDGLSEKHPHFK-----NVKE 233
DB 102 VHAYAYQSVPIGVSQ--IKAPALHSQGITGSNWKVAVIDSGIDSSHPDLKVAGGASWPS 160

QY 234 RFTNWTNERTLDDGLGHGTFVAGVIASMRE---CGFPADAEHLHFRVFTNNQVSYTWNFL 290
DB 161 ETN-----PFQDNNSHGTHVAGTVAALNNSIGVLGVAPASASLYAKVGLGADSGOYSMII 215

QY 291 DAFNYAILKIDVNLNLSIGGPDFMD--HPFDVKWELTANNVIMVSAIGNDGP--LYGTL 346
DB 216 NGIEWAANNNDVIMNLSLGGPSGAALKAAYDKA---VASGVVVAAGNENSTSGSSTV 272

QY 347 NNPAQMDQMDVIGVGGIDFEDNIAFRFSRGMTTWELPGCYGRMKPDIVTYGAGVSGS--GVK 404
DB 273 GYPGKYPSTAVGAVDSSNQRAFSSVSGP---EL-----DVMAPGVSTQSTLPGNK 320

QY 405 GGCRAISGTSVASPVVAGAVTLLVSTVQKRELVPASMKQALIASARRLPGVNMFEQGHG 464
DB 321 YG--AYNGTSMASPHVAGAAALILS---KHPNWTNTQVRSSLENTTTKLG--DSFYFKGK 373

QY 465 KLDLIRAVQ 473
DB 374 LINVQAAQ 382

RESULT 11


```

QY 453 LFGVNMFGQHGKLDL 468
DB 297 -----AQSGLLKL 305

RESULT 14
C41335
microbial serine proteinase (EC 3.4.21.-), intracellular - Bacillus polymyxa
C:Species: Bacillus polymyxa
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: C41335
J:Takekawa, S.; Uozumi, N.; Tsukagoshi, N.; Ueda, S.
R: Bacteriol. 173, 6820-6825, 1991
A:Title: Proteases involved in generation of beta- and alpha-amylases from a large amylo
A:Reference number: A41335; MUID:92041565; PMID:1834632
C:Accession: C41335
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <TAK>
A:Cross-references: GB:D00862; NID:g216285; PID:BAA00735.1; PID:g216286
C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:40-236/Domain: subtilisin homology <SBT>
F:49,86,244/Active site: Asp, His, Ser #status predicted

Query Match 5.3%; Score 299; DB 1; Length 326;
Best Local Similarity 30.3%; Pred. No. 1.3e-13;
Matches 94; Conservative 54; Mismatches 126; Indels 36; Gaps 12;

QY 189 IPRVQATLQADVLQMGYTGANVRVAVFDTLGSEKHPHF-K-NYKRTNWTNERTID--- 244
DB 22 IPRGV-EMIQAPAVNQ-TRGRGVKVAVLDTGCDADHPDKARIIGRNTDDGDEPEI 79

QY 245 --DLGHTGVAGVIASMR--CQGEAPDAELHIFRVTNNQVSYTSWFLDAFNAILK 299
DB 80 FKDYNGHTHVAGTIAATENENGVVGVAPEADLLIKVLNKGSGQYDWIIQGIYAT 139

QY 300 KIDVLNLSIGDPMDHPV-DKWELTANNVMSAIGNDQ---PLYGTLNAPADQMDV 355
DB 140 KVDILNLSIGPE--DVPHEAKKAVASQILVWCAGNEGDDRTDELGYPCYNEV 197

QY 356 IQVGIDFENIARFSSRCMTWELPGYGRMKPDIVYAGVGRSGVKGCCRALSGTSV 415
DB 198 ISVGAINFDRHSEFSN-----SNNEVDLVAPGEDILSTVPGKATYSGTSM 245

QY 416 ASPVYAGAVTL---LVSIVQKRELVPASMKQALIASRLP-GVNMFEQGHGKLDLRA 471
DB 246 APHVALGALALIKOLANASFENDUTEPELYAQLI---RTIPLGNSPKMEGNLLYLAV 302

QY 472 YQILNSYKPK 481
DB 303 EELSRIFDTQ 312

RESULT 15
JC1085
subtilisin (EC 3.4.21.62) precursor - Bacillus licheniformis
N:Alternate names: alkaline proteinase
C:Species: Bacillus licheniformis
C:Date: 09-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 31-Mar-1997
C:Accession: JC1085
R:Lei, H.; Hong, Y.; Zhang, Y.Y.; Shen, T.J.
Chinese Biochem. J. 9, 441-447, 1993
A:Title: PCR amplifying cloning and sequencing of the coding sequence of the alkaline p
A:Reference number: JC1085
A:Accession: JC1085
A:Molecule type: DNA
A:Residues: 1-275 <LEI>
A:Note: The translation of the start codon ATG is not given in this paper
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:24-235/Domain: subtilisin homology <SBT>
F:33,64,221/Active site: Asp, His, Ser #status predicted

```

```

Query Match 5.3%; Score 295; DB 2; Length 275;
Best Local Similarity 31.5%; Pred. No. 2e-13;
Matches 93; Conservative 46; Mismatches 122; Indels 34; Gaps 9;

QY 193 VAQT-----LQADVLQMGYTGANVRVAVFDTLGSEKHPHFKNKERTNWTNERTLDD 245
DB 1 MAOTVPYGIPLIKRADKVOAGGKGANVRVAVLDTGICQASHPDNLNVVVGASFVAGEAYNTD 60

QY 246 GIGHTFTFVAGVIASMR--CQGEAPDAELHIFRVTNNQVSYTSWFLDAFNAILKKID 302
DB 61 GNGHGTHTVAGTIAALDNTFVLGVAPSVSLYAVLVNLSGSGSYSGIVSGIEWATTNGMD 120

QY 303 VLNLSIGDPMD--HPFDVKWELTANNVMSAIGNDQPL--YGTLNAPADQMDVIGV 358
DB 121 VIMSLSGSGSTAMQAVDNAY---ARGVVVAAGNSGSGNTNTIGYPAKVDSTVAV 177

QY 359 GQIDFENIARFSSRCMTWELPGYGRMKPDIVYAGVGRSGVKGCCRALSGTSVASP 418
DB 178 GAVDSNRSRASFSSVGA---EL-----EVNAPAGVISTYPTNTYATLNGTSMASE 225

QY 419 VVAGAVTLVSTVQKRELVPASMKQALIASRLPQVNMFEQGHGKLDLRAVQ 473
DB 226 HVAGAAALILS---KHPNLSASQVRNRLSSTATYLG--SSFYVGKGLINVEAAQA 275

RESULT 16
JC5460
intracellular alkaline serine proteinase (EC 3.4.-.-) - Thermoactinomyces sp.
C:Species: Thermoactinomyces sp.
C:Date: 17-Jun-1997 #sequence_revision 17-Sep-1997 #text_change 20-Jun-2000
C:Accession: JC5460; PC4328
R:Tsuchiya, K.; Ikeda, I.; Tsuchiya, T.; Kimura, T.
A:Title: Cloning and expression of an intracellular alkaline protease gene from alkal
A:Reference number: JC5460; MUID:97212029; PMID:9058969
A:Accession: JC5460
A:Molecule type: DNA
A:Residues: 1-321 <TSU>
A:Cross-references: DBRI:D87557; NID:g1536872; PID:BAA13418.1; PID:g1536873
A:Experimental source: strain HS682
A:Accession: PC4328
A:Molecule type: protein
A:Residues: 16-45 <TSU2>
C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology
C:Keywords: hydrolase
F:16-321/Product: intracellular alkaline serine proteinase #status predicted <MAT>
F:40-263/Domain: subtilisin homology <SBT>
F:32,86,220/Active site: Asp, His, Ser #status predicted

Query Match 5.2%; Score 294; DB 2; Length 321;
Best Local Similarity 27.7%; Pred. No. 3e-13;
Matches 91; Conservative 54; Mismatches 111; Indels 72; Gaps 12;

QY 186 LRAIPROVA-----QTLQADVLQMGYTGANVRVAVFDTLGSEKHPHFKNVK 232
DB 4 VRLIPYEVTSIQDDTKKIPPTGEMTEAPDLWOQYKGGIVVAVLDTGCDVEH---YELR 60

QY 233 ERTNWTNERTLDG-----LGHTFVAGVIASMR--CQGEAPDAELHIFRVTNN 281
DB 61 DRIIGKHNVTSDGNDPELVSDQNGHGTHTVCGTIAATENDRAIGVAPCQLLVKVLNSR 120

QY 282 QVSYTSWFLDAFNAIL-----KIDVLNLSIGDPMDHPFDVKWELTANNVMSA 335
DB 121 GFGTEWVVGIRHAIWNEGPNGERKQVQLSLSGCKE--NDPLRLHAIKENVASGRLLVCA 179

QY 336 IGDGPLYGLTNAPADQ-----DVTGVGIDFEDNTARFSSRCMTWELPGYGRMK 388
DB 180 AGND-----GGNSETDEFAYPGAYPEVQVQVSGVSLSGEISRFSN-----SNCK 223

QY 389 PDIVYAGVGRSGVKGCCRALSGTSVASPVVAGAVTLVSTVQK---RELVPASMKQA 445
DB 224 IDLVAPGEKILSTPGDKFATLTGTSMTAPHTVGAALLTERFEREFERKTEPELPAQL 283

```

A::Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A::Reference number: A83650; MUID:20512582; PMID:11058132

A::Accession: G83922

A::Status: preliminary

A::Molecule type: DNA

A::Residues: 1-322 <STO>

A::Cross-references: GB:AP001514; GB:BA000004; NID:g10174613; PIDN:BAB05902.1; GSFPDB:GNOC

A::Experimental source: strain C-125

C:Genetics:

A::Gene: isp

C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology

Query Match 5.2%; Score 293.5; DB 2; Length 322;
Best Local Similarity 27.9%; Pred. No. 3.2e-13;
Matches 91; Conservative 55; Mismatches 113; Indels 67; Gaps 12;

Qy 186 LRAIPRQA-----QTLAADVLWQMGTGANVRVAVFOTGLSEKHPFKN-V 231
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :
Db 4 VRLIPEVTSIQDTHKKPIPGMIEMRAPDLWOOGYKGIGIVAVLDTGCDVEHYELDR I 63
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Qy 232 KERTNWNE-----RTLDDLGHGHTFVGAGVIASM---RECQGPAFDAELHIFRVFTNNOV 283
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Db 64 IGKHNTVEDNGPDEIVSDQHGHCTHCVTATANDRGVIGVAPECQLLVKVLSNRGF 123
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Qy 284 SYTSWFLDAFNAIL-----KKIDLNLISGGDFMDPHFYDKVWELTFANNVMYSALG 337
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Db 124 GTTEWVEGIRHAINEWGPNKEGVQVSLMSLGSKE--NDPRLHDIAIKEAVASGRLLVCACAG 182
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Qy 338 NDGPPLYGTLNNPADQM-----DVTVSGGDIEDNIARESSRCMTWTWELPCGYGERMKPD 390
 || | | | | | | | | | | | | | | | | | | :| :| :

Db 183 ND----GDGNETDEFAYFCAYEVVVQVSUSGEISRFSN-----SNCKID 226
 || | | | | | | | | | | | | | | | | | | :| :| :

Qy 391 IVTYGAGVRGSKGCCRALSGTFSVASPVVAGAVTTLLSVTKQ---RELVPNSMKQAL- 446
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Db 227 LVAPGEKILSTYPDKNFAITLTGMTATPHVTGAALLIEKFEREFEKITEPELFACLAK 286
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

Qy 447 -----IASARKLPVNWFEGHGKL DL 468
 :: :| :| | | | | | | | | | | | | | | :| :| :

Db 287 RTVSLSYSRKL-----CGNGLLKL 305
 :| | |:| :| :| | | | | | | | | | | | | | :| :| :

A:Accession: F69646
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-70, T, 72-159, 'K', 161-319 <UN>
A:Cross-references: GB:299110; GB:AL009126; NID:G2633472; PIDN:CAB13176.1; PID:G26336136
A:Experimental source: strain 168
C:Genetics:
A:Gene: ispA
C:Superfamily: Bacillus intracellular serine proteinase; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:41-260/Domain: subtilisin homology <SBT>
F:50,87,246/Active site: Asp, His, Ser #status predicted

Query Match 5.1%; Score 289; DB 2; Length 319;
Best Local Similarity 32.0%; Pred. No. 6.7e-13;
Matches 80; Conservative 45; Mismatches 95; Indels 30; Gaps 9;

Qy 195 QTLQADVLMQGYGTANRVAVFDTLGLSEKHPFKN-VKERTNWTN-----ERTLDDGLG 248
Db 27 KVTAPEMWAKGVKKNIKVAVLDTGDTSHPLDKNOITGGKNFSDDDGKEDAISYNG 86
Qy 249 HGTFVAGVIA---SMRECGQFAPDAELHIFRVF--TNNQVSYTSWFLDAFNAYILKKIDV 303
Db 87 HGTHVAGTTAANDSNGGIAGVAPEASLLIVKVLGGENGSGQY-EWIIINGINYAVEOKVDI 145
Qy 304 LNLISGSGPMDHPFDVK-TWELTANNVMSAIGNDG---PLYGTLNNPADQMDVIGVG 359
Db 146 ISMSLGPS--DYPELEEAQVNAKVGVLVYVCAAGNEGDDGDERTEELSYPAAYNEVIANG 203
Qy 360 GIDFEDNIARFSRGMGTTWELPGGYGRMKPDIVTYGAGVRGSGVKGCRLSTGTSVASPV 419
Db 204 SVSVAARELSFSN-----ANKEDILVAPGENILSTLPNNKYKILGTGTSMAAPH 251

pyrolysin (EC 3.4.-.-) - Pyrococcus furiosus
C;Species: Pyrococcus furiosus
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 24-Oct-2000
C;Accession: T28159
R;Voorhorst, W.G.B.; Eggen, R.I.L.; Geerling, A.C.M.; Platteuw, C.; Slezzen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
A;Title: Isolation and characterization of the hyperthermostable serine protease, pyrolysin
A;Reference number: Z20481; MUID:96355370; PMID:8702780
A;Accession: T28159
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1398 <VOO>
A;Cross-references: EMBL:U55835; NID:g1556462; PID:g1556463; PIDN:AAB09761.1
A;Experimental source: DSM3638
C;Genetics:
A;Gene: pls
C;Keywords: hydrolase; serine proteinase

Query Match 5.0%; Score 281.5; DB 2; Length 1398;
Best Local Similarity 22.9%; Pred. No. 2.5e-11;
Matches 164; Conservative 84; Mismatches 222; Indels 245; Gaps 34;
QY 245 DGLGHGTFVAGVTA-----SMRECOGFAPDAE 271
DB 361 DGHGHGTHVAGVAGYDSNDADWLSMSEGEVFSRLYGDYTNVTTDVGAPGAQ 420
QY 272 LHFRFVETNNQVSYTWFDAFNAILKIDVLNLSIGGDFMDHPVD-----KWVE 324
DB 421 IMAIRVLRSDGRSMWDIIIEGMYAATHGADVLSMLGG---NAPYLDGTDESVADE 476
QY 325 LTAN-NVIMVSAIGNDGLYGTLLNNPADQMDVIGV-----GIDFED----- 365
DB 477 LTERKGVVFWVIAAGNEPGNIGVSGVATKALTGTGAAAVPINVVYVSQALGPDYGF 536
QY 366 -----NIARFSSRGMTTWELPGCYGRMKPDIVTYGAGVRSYGVK---GGCRALSGTS 414
DB 537 YFPATNVRIAFSSRG-----PRIDGEIKPNVAPGYGYSLLPMWIGGADFMSTGS 590
QY 415 VASPVVAGATLLVSTVQREL-VNPSMKOALIASARRLP-----VNMEEQHGKGL 466
DB 591 MATPHSVGVVALLISGPKPGIYNDPIIKKVLSEGTATWLEGGDYTKQKYTELDQHGVL 650
QY 467 DLRAVQILNSYKQASLSYSIDLTECPYMWYCSQPIYGMPTVVNVTILNGMGTG 526
DB 651 NVTKSWEILKA-----INGTLPVDEHWADKS--YSDFAEYLGVDVIRGLYARN 697
QY 527 RYVDKPDWQYLPQNGDNIETVAFSYSSVLPWPSYLAISVTKKAASW-EGTAQGHVMI 585
DB 698 SIPDIVEW--HIKYVD-----TEYTFEYATE---PWIKPFVSGSVIL 737
QY 586 TVASPAETESKGAETSTVKLPK-----VKII---PTPPSRKRVLDQVHNLRY 634
DB 738 -----ENTEFVLRYKYDVEGLEPGLYVGRILIDPTTP-----VIEDELLNTIVI 783
QY 635 PGTFPRDLNRMKNLDLW-----NGDHITNFRDMYQHLRSMGYFVEVLGAPFTCFDASQY- 690
DB 784 PERFTPEN-----NVTLTWYDINGPEMVTH-----HFTVPEGVGVLYAMTYWDYGLYR 833
QY 691 --GTLMLVDSEEEYFPEETIAKLRRVDNGLSVLIFSDWYNTVYMRKVFTDENTROWMP 748
DB 834 PDGMFV-----FPQDYLPAAVSNMP-----GNW----- 859
QY 749 DTGGANIPALNELLSVNMGMFSDG-LYEGEFTLANHDMYYVAGCSIAKFPEDGVVIT--- 804
DB 860 -----EL--VWT-GFNAPLYESGFLVRIH-----GVEITPSV 889
QY 805 -----QTFKQGLEVLKQETAVVENVPI-----LGLYQIPAE--GGGRIVLYG 845
DB 890 WYINRYLDNTFEESIEFNITNIYAPINATFLIPGLGTYNASVESVGDGEFFTKG 944

RESULT 23
JC4802

alkaline proteinase (EC 3.4.21.-) precursor - Thermoactinomyces sp. (strain E79)
C;Species: Thermoactinomyces sp.
A;Variety: strain E79
C;Date: 15-Aug-1996 #sequence_revision 15-Oct-1996 #text_change 09-Jun-2000
C;Accession: JC4802
R;Lee, J.K.; Kim, Y.O.; Kim, H.K.; Park, Y.S.; Oh, T.K. Biosci. Biotechnol. Biochem. 60, 840-846, 1996
A;Title: Purification and characterization of a thermostable alkaline protease from T
A;Reference number: JC4802; MUID:96261070; PMID:8704314
A;Accession: JC4802
A;Molecule type: DNA
A;Residues: 1-384 <LEE>
A;Cross-references: GB:U31759; NID:g1389689; PIDN:AAB36499.1; PID:g1683629
A;Experimental source: strain E79
C;Comment: This protein is thermostable.
C;Function:
A;Description: extracellular alkaline serine proteinase [validated, MUID:96261070]
C;Superfamily: subtilisin; subtilisin homology
C;Keywords: hydrolase; serine proteinase
F;1-25/Domain: signal sequence #status predicted <SIG>
F;26-106/Domain: propeptide #status predicted <PRO>
F;107-383/Product: alkaline proteinase #status experimental <MAT>
F;134-344/Domain: subtilisin homology <SBT>
F;143,176,330/Active site: Asp, His, Ser #status predicted

Query Match 4.9%; Score 276.5; DB 2; Length 384;
Best Local Similarity 24.0%; Pred. No. 7.1e-12;
Matches 115; Conservative 75; Mismatches 171; Indels 119; Gaps 20;

QY 9 LLLVLLCGKKHGLDRLEKSKFEKAPCGGSHLTLKVEFSSTVVEY---EYVAFNGYFT 65
DB 11 LLLVLLLA-----VPG-----TMFAASPASTDYDYPGELIVFKDGIS 48
QY 66 AKARNSFISS-ALKSSEVDNWRRIIPRNPSSDYSPDFEVIQIKKQAGLLTLDEHPNIK 124
DB 49 AQSTQSTHAQYGAKSIE-----KSKY-LGFEVVKFDGSEVMKIEKYKNNPNVE 95
QY 125 RVTQPRKFRSLKYAESDPTVPCNETRWSOKWSRPLRSLASLSIGSGFWHATGRHSRR 184
DB 96 YVEPNHYV--HIMTTPNDLT-----SRQWG----- 118
QY 185 LLRAIPROVAQTQADVLWQMGYTGANRVAVFTGLSEKHPH--KNVKERTWNTWERT 242
DB 119 -----PQVQAPQAWDVTRSSSTVAIVDTGVQTNHPLDQLQGVQYDVFVDNSN 169
QY 243 LDGSLGHGTFFVAGVIASMR-----CQGFAPDAELHIFRVTTNNQVSYTSFELDAFNAIL 298
DB 170 PQDNGHGTCHAGTAAAVTNNGTGACMAPNASIMPVRLNNSGSGTMAAVANGIAVAAQ 229
QY 299 KKIDVLNLSIGGPDFMD--HPFVDKVMWELTANNVIMVSAIGNDGLYGTLLNNPADQMDVI 356
DB 230 NGADVLSLGGTSGSALQSAVQOANN---SGAVVVAAGNSS--STPNYPAYISOAI 284
QY 357 GVGIDFEDNIARFSSRGMTTWELPGCYGRMKPDIVTYGAGVRSYGVKGCRAISGTSVA 416
DB 285 AVASTDSNDSLSYFSNYG--SW-----VDVAAPGSGNIYTYLNSSVASLSGTSMA 332
QY 417 SPVVAGAVTLLVSTVQRELVPNSPMKQALIASARRLPVGNMFFQGHGKLDLLRAYQILN 476
DB 333 TPHVAGIALLAS--OGR-----SNSQIRAAIENTADKISGTGYTFQ-HGRIN---AYKAVN 383

RESULT 24
G95392

probable protease [imported] - Sinorhizobium meliloti (strain 1021) magaplasmid pSyna
C;Species: Sinorhizobium meliloti
C;Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C;Accession: G95392
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli
A;Reference number: A95262; MUID:21396509; PMID:11481432

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: June 2, 2003, 02:50:54 ; Search time 113 seconds
(without alignments)
2855.079 Million cell updates/sec

Title: US-09-830-837-6
Perfect score: 5617
Sequence: 1 MKLVNIWLLVLLCGKKH.....PRVKRPLMQVHPKTPSV 1052

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DRV=xlh
-O=/cgn2_1/USPTO.spool/US09830837/runat_23052003_181925_8091/app_query.fasta.1.1223
-DB-Issued_Patents_NA_QFMT=fastp -SUFFIX=p2n.rni -MINMATCH=0.1 -DOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837.ecgn.1.1.46.runat_23052003_181925_8091 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5617	100.0	4338	US-09-360-237-4	Sequence 4, Appli
2	5487	97.7	4198	US-09-360-237-2	Sequence 2, Appli
3	338.5	6.0	1500	US-08-322-965-1	Sequence 1, Appli
4	331	5.9	1497	US-08-322-677A-6	Sequence 6, Appli
5	331	5.9	1497	US-08-322-676-6	Sequence 6, Appli
6	331	5.9	1497	US-08-898-218-6	Sequence 6, Appli
7	331	5.9	1497	US-08-848-793-6	Sequence 6, Appli
8	331	5.9	1497	US-09-445-270-1	Sequence 1, Appli
9	331	5.9	1513	US-09-178-155-1	Sequence 1, Appli
10	330	5.9	2809	US-09-000-016-1	Sequence 1, Appli
11	330	5.9	2809	US-09-514-340-1	Sequence 1, Appli
12	329	5.9	1496	RE34606-1	Patent No. RE34,60

13	328.5	5.8	1494	4	US-09-255-502-1	Sequence 1, Appli
14	328.5	5.8	1497	6	5472855-1	Patent No. 5472855
15	328.5	5.8	1868	1	US-08-069-863-1	Sequence 1, Appli
16	328.5	5.8	1868	1	US-08-309-069-1	Sequence 1, Appli
17	328.5	5.8	1868	5	PCT-US95-05520-1	Sequence 1, Appli
18	328.5	5.8	1868	5	US-08-460-343B-1	Sequence 1, Appli
19	328.5	5.8	8119	1	US-08-398-028B-1	Sequence 1, Appli
20	328.5	5.8	8119	1	US-08-504-265B-1	Sequence 1, Appli
21	328.5	5.8	8119	2	RE34606-4	Patent No. RE34,60
22	325.5	5.8	1498	3	US-09-023-173-4	Sequence 4, Appli
23	325.5	5.8	1452	1	US-08-397-602A-8	Sequence 8, Appli
24	324	5.8	2539	3	US-09-000-016-3	Sequence 3, Appli
25	324	5.8	2539	3	US-09-514-340-3	Sequence 3, Appli
26	324	5.8	2539	4	US-08-894-818B-6	Sequence 6, Appli
27	320.5	5.7	1977	4	US-08-894-818B-2	Sequence 2, Appli
28	320	5.7	1977	4	US-09-445-472-11	Sequence 11, Appli
29	320	5.7	1977	4	US-08-504-265B-74	Sequence 74, Appli
30	319.5	5.7	1146	2	US-08-685-774-1	Sequence 1, Appli
31	319	5.7	1457	1	US-09-445-472-2	Sequence 2, Appli
32	302	5.4	1236	4	US-08-894-818B-4	Sequence 4, Appli
33	302	5.4	1566	4	US-08-894-818B-34	Sequence 34, Appli
34	302	5.4	1962	4	US-09-445-472-15	Sequence 15, Appli
35	302	5.4	1962	4	US-07-772-087-1	Sequence 8, Appli
36	288	5.1	1867	1	US-08-750-532-8	Sequence 7, Appli
37	285	5.1	4765	4	US-09-445-472-5	Sequence 5, Appli
38	285	5.1	4765	4	US-08-875-154-1	Sequence 1, Appli
39	285	5.1	10216	2	US-08-269-050-1	Sequence 1, Appli
40	284	5.0	1143	6	5336611-1	Patent No. 5336611
41	283	5.0	1143	6	5217878-1	Patent No. 5217878
42	283	5.0	1320	6	US-09-049-867-1	Sequence 1, Appli
43	283	5.0	1350	3	US-08-322-677A-11	Sequence 11, Appli
44	283	5.0	1140	1		
45	282	5.0				

ALIGNMENTS

RESULT 1
US-09-360-237-4
; Sequence 4, Application US/09360237
; Patent No. 6322962
; GENERAL INFORMATION:
; APPLICANT: BROWN, MICHAEL S.
; APPLICANT: CHENG, DONG
; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JURO
; TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; FILE REFERENCE: UTXD:567
; CURRENT APPLICATION NUMBER: US/09/360,237
; CURRENT FILING DATE: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096,571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Human
US-09-360-237-4

Alignment Scores:
Pred. No.: 0
Score: 5617.00
Length: 4338
Matches: 1052
Conservative: 0
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Mismatch: 0
Query Match: 100.00%
Indels: 0
Gaps: 0
DB: 4

US-09-830-837-6 (1-1052) x US-09-360-237-4 (1-4338)

Db 2657 GTCATCTTCAGTGGTACAACTCTCTGTATGAGAAAAGTGAAGTTTATGATGAA 2716
QY 741 AsnThrArgGlnTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2717 AACACAGGAGTGGTGGATCGCGATACCGGAGAGCTAACATCCAGCTCTGAATGAG 2776
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTrpGluGlyPheThrLeu 780
Db 2777 CTGCTCTCTGTGGAACTGGGTTCCAGGATGGCTGTATGAAGGGAGTTCCACCTG 2836
QY 781 AlaAsnHisAspMetTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2837 GCCAACCATGACATGATTATGCGTCAGGTCAGGTCAGCATCCGGAAGTTTCCAGAAGATGGC 2896
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2897 GTGCTGATACACAGACTTTCAGGACCAAGGATGAGGTTTAAAGCAGGAAACAGCA 2956
QY 821 ValValGluAsnValProIleLeuGlyLeuTrpGlnIleProAlaGluGlyGlyArg 840
Db 2957 GTTGTGAAAACGTCCCATTTTGGGACATTATACATCCAGCTCAGGTCGAGCGCG 3016
QY 841 IleValLeuTrpGlyAspSerAsnCysLeuAspSerHisArgGlnLysAspCysPhe 860
Db 3017 ATTGTACTGTATGGGACTCCCAATTCGTTGGATGACAGTCACCGACAGAAGGACTGCTT 3076
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTrpThrSerTrpGlyValThrProProSerLeuSer 880
Db 3077 TGGCTTCTGGATGCCCTCTCCAGTACACATCGTATGGGTGACCGCTAGCCCTCAGT 3136
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3137 CACTCTGGAAACCGCCAGCGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG 3196
QY 901 GluGlyAsnHisLeuHisArgTrpSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3197 GAAGGAAACCACTTCTCATCGGTACTCCAAAGGTTCTGGAGGCCCATTTGGGAGACCCAAA 3256
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3257 CCTCGGCTCTACACGCTGTCCACGCTGTCTTGGGCCAGGCCAGGCTTTAAACGAG 3316
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3317 ACGGCGCCCAAGAACCTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGGTG 3376
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyLys 980
Db 3377 GTGTTACCCCACTTTCGATCGAATCGCCCTCAAGTGAAGGCCCTTGTCCCTGGAGAGC 3436
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTrpAsnGlnGluValGlyGln 1000
Db 3437 GCGGCTGGGACATTCCTGGGGATCATGCTGGCGGCTACACAGGAGGTGGGCCAG 3496
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3497 ACCATTCTCTCTTTCCTCTCTGGGAGCATGTGCTCGCTCTCTTGTGGTACAA 3556
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3557 ATCAACAGGCCCAAGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3616
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3617 ATGCACAGGTTTACCCCGCCCAAGACCCCTCTGGTG 3652

RESULT 2

US-09-360-237-2

; Sequence 2, Application US/09360237

; Patent No. 6322962

; GENERAL INFORMATION:

; APPLICANT: BROWN, MICHAEL S.

; APPLICANT: CHENG, DONG

; APPLICANT: ESPENSHADE, PETER J.
; APPLICANT: GOLDSTEIN, JOSEPH L.
; APPLICANT: RAWSON, ROBERT B.
; APPLICANT: SAKAI, JURO
; TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF
; FILE REFERENCE: US/09-360, 237
; CURRENT APPLICATION NUMBER: 1999-07-23
; CURRENT FILING DATE: 1999-07-23
; EARLIER APPLICATION NUMBER: 60/096, 571
; EARLIER FILING DATE: 1998-08-14
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 4198
; TYPE: DNA
; ORGANISM: Cricetulus griseus
; US-09-360-237-2

Alignment Scores:

Pred. No.: 0 Length: 4198
Score: 5487.00 Matches: 1021
Percent Similarity: 98.57% Conservative: 16
Best Local Similarity: 97.05% Mismatches: 15
Query Match: 97.69% Indels: 0
DB: 4 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-360-237-2 (1-4198)

QY 1 MetLysLeuValAsnIleTrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
Db 387 ATGAAGCTCATCAACATCTGGCTTCTTCTGCTGGTGGTTTCTCTGTGGAAAGAGCAT 446
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
Db 447 CTGGGTGACAGGCTGGGGAAGAAAGCTTTGAAAGGCATCATGCTCAGCTTCTCCAC 506
QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTrpGluTrpIleValAlaPhe 60
Db 507 CTGACTTTGAAGTGGAAATCTCTCAACTGTGTGGATATGAATATATTGTGGCTTTC 566
QY 61 AsnGlyTrpPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSer 80
Db 567 AACGGATACTTACAGCAAGCTAGAACTCATTTATTTCAAAGTGTCTGAAAGCAGT 626
QY 81 GluValAspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTrpProSerAsp 100
Db 627 GAAGTAGACAACCTGGAGAATTATACCTCGGAACAACCCATCCAGTACTACCTAGTAT 686
QY 101 PheGluValIleGlnLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
Db 687 TTTGAGTGTATTCAGATAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCAT 746
QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTrpAlaGlu 140
Db 747 CCAACATCAAGCGGGTGACACCTCAACGCAAGTCTTTCTGCTTGAAGTTTGTCTGAA 806
QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
Db 807 TCTGACCCCATTTGTCATGTAATGAACCTGGTGGAGCAGAGAGTGGCAGTCAACCA 866
QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
Db 867 CCCTCAGATCAT 926
QY 181 SerSerArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
Db 927 TCAAGCGCGGATTTGCTGAGAGCATTCCTCGACAGGTTGCCAGACATTCGAGGAGAT 986
QY 201 ValLeuTrpGlnMetGlyTrpThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
Db 987 GTCTGTGGCAGATGGGATACAGAGTGTATGTCTAGGTTGTCTGCTGCTGCTGCTGCTGCT 1046

Db 3207 ACTGCCGCCAGTAATCTTTGGAAACATCAGAGCTGCTCTCCATTCACCTGGAGCAAGTA 3266
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCCAACTTCATCGAATCGCCCTCAAGTGAGACCTTTGTCCCTGGAGAAAGT 3326
QY 981 GYAlaTrpAspIleProGlyGlyIleMetProGlyArgTrpAsnGlnGluValGlyGln 1000
Db 3327 GTGCGCTGGACATCTCTGGAGGATCATGCTGCGCGCTACAAACAGAGGTGGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3387 ACCATCCCTGCTCTTTCCTCTCCCGGAGCCATGTCGCGCTCTTCTTGTGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGCCCAAGCGCGCGAGGAGGAGGCCAGGCCAAGCGCTCCACAGCTT 3506
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCCACCACCAAGGACCCCGTCAGTG 3542

RESULT 3

US-09-832-965-1
; Sequence 1, Application US/08322965
; Patent No. 5733473
; GENERAL INFORMATION:
; APPLICANT: Johnston, James P.
; APPLICANT: Lenoir, Pierre M.A.
; APPLICANT: Thoen, Christian A.J.K.
; TITLE OF INVENTION: Liquid Detergent Composition Containing
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 5299 Spring Grove Avenue
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45217
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,965
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 252
; ATTORNEY/AGENT INFORMATION:
; NAME: Allen, George W.
; REGISTRATION NUMBER: 26,143
; REFERENCE/DOCKET NUMBER: CM-352MC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 513/627-5946
; TELEFAX: 513/627-8118
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1500 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 455..1282
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 137..1282
US-08-322-965-1

Alignment Scores: 1.53e-23 Length: 1500
Pred. No.:

Score: 338.50 Matches: 133
Percent Similarity: 40.58% Conservative: 76
Best Local Similarity: 25.83% Mismatches: 180
Query Match: 6.03% Indels: 126
DB: 1 Gaps: 21
US-09-830-837-6 (1-1052) x US-08-322-965-1 (1-1500)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
Db 233 AGCAGTACA-----GAAAGAAATATCATTTGCGGATTT-----AAA 268
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
Db 269 CAGACATAGTCCCATGAGTTCGCCAAGAAAAGATGTTAT----- 313
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
Db 313 ----- 313
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
Db 314 ---TCGAAAAGCGCGAAGGTCAAAAGCAATTTAAGTATGTTAACGCGCGCAGCA 370
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
Db 371 ACATTGGATCAAAAAGCTGTAAAGAAATTGAAA-----AAAGATCCGACGGTGA 421
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSer 166
Db 421 ----- 421
QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
Db 422 -----TGTGGAAGAACATCATATGCACATGAATATGCG 457
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
Db 458 CAATCTGTCTTATGGCATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 514
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
Db 515 TACACAGGCTCTAACGTAAAGTAGCTGTATGCACAGCGGAATGACTCTTCTCATCCT 574
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
Db 575 GACTTA---AACGTACAGCGCGGAGCAAGCTTCGTACTCTCTGAAACAAACCCATACCAG 631
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 632 GACGGCAGTCTCACGGTACGGCATGTAGCGGTACGATGCGGCTCTTAATACTCAATC 691
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
Db 692 GGTGTTCTGGCGGTAGCCCAAGCGCATCATATATGCAGTAAAGTGTGTGATCAACA 751
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
Db 752 GGAAGCGGCAATATAGCTGATTTAACGGCATTCAGTGGGCCATTTCCCAACAATATG 811
QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
Db 812 GATGTTATCAACATGAGCCTTGGCGACCTACTGTTCTACACGCGCTGAAACAGTCGT 871
QY 320 AspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 872 GACAAAGCC-----GTTCCACGCGGTATGCTGCTGCTGCCGCGGAAACGAA 922
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnProAlaAspGlnMetAspValIleGly 357
Db 923 GGTTCATCCGGAAGCACAAAGCAGTCGCGCTACCTGCGCAAAATATCTTCTACTATGCA 982
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377

```

||||| :||||| :||| :||||| :|||
983 GTAGTGGTAAACAGCAGCAACCAAGAGCTTCAATTCACGCGAGGTCT----- 1036
QY 378 TrpGluLeuProGlyGlyTyrglyArgMetLysProAspIleValThrTyrglyAlaGly 397
||||| :||||| :||| :||||| :|||
1037 ---GAGCTT-----GATGTGATGGCTCCTGCGGTGCC 1066
QY 398 ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaLaser 417
||||| :||||| :||| :||||| :|||
1067 ATCCAACAGCAGCTCTCTGGAGGCACTTACGGCGCTTATACGGAACGCTCATGGCGACT 1126
QY 418 ProValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
||||| :||||| :||| :||||| :|||
1127 CTTCAAGTTCGGAGCAGCAGCGTAAATCTTCTC-----AAGCACCAGCTTGG 1177
QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
||||| :||||| :||| :||||| :|||
1178 ACAACAGCGCAGCTCTCTGGAGGCACTTACGGCGCTTATACGGAACGCTCATGGCGACT 1231
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuArgAla-TyrGlnIleLeuAsnSe 477
||||| :||||| :||| :||||| :|||
1232 TCTTCTACTATGAAAGGGTAAATCAACGTACAGCAGCTGCACAA-----AATAG 1285
QY 477 rTyLysProGlnAlaSerLeuSerProSerTyrlleAspLeuThrGluCysProTyrMe 497
||||| :||||| :||| :||||| :|||
1286 TAAAGAGCAGCGT-----CCTCATACCTGCTCTCTTTTATTGTCAGCAT-- 1334
QY 497 tTrpProTyrCysSerGlnProIleTyrglyGlyMetProThrValValAsnValTh 517
||||| :||||| :||| :||||| :|||
1335 -----CCTGATGCTCCGGCGCATTC 1354
QY 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgI 528
||||| :||||| :||| :||||| :|||
1355 TCTTCTTCTCCGATGTTGAATCGTTCATGATGATGCGGATGCTGCTCTGAAATC 1414
QY 528 eValAspLysProAspTrpGlnProTyrLeuProGlnAsnGly 542
||||| :||||| :||| :||||| :|||
1415 TTCACAGCAGCGAGATCACTGCTCAGCCCGCTCAGCGC 1457

```

RESULT 4

```

US-08-322-677A-6
; Sequence 6 Application US/08322677A
; Patent No. 567272
;
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Chanchal K.
; APPLICANT: Burns, Michael E.
; APPLICANT: DiGiulio, David N.
; APPLICANT: Getty, Edward E.
; APPLICANT: Hartshorn, Richard T.
; APPLICANT: Willey, Alan D.
; APPLICANT: Brode, Philip F., III
; APPLICANT: Barnett, Bobby L.
; APPLICANT: Rubingh, Donn N.
; TITLE OF INVENTION: Bleaching Compositions Comprising Protease Enzymes
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: The Procter & Gamble Company
; STREET: 11810 East River Road
; CITY: Cincinnati
; STATE: OH
; COUNTRY: USA
; ZIP: 45253-8707
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322.677A
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Zerby, Kim William

```

```

;
; REGISTRATION NUMBER: 32,323
; REFERENCE/DOCKET NUMBER: 5041R
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 627-2885
; TELEFAX: (513) 627-0318
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1497 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-322-677A-6
;
; Alignment Scores:
; Pred. No.: 8,39e-23 Length: 1497
; Score: 331.00 Matches: 124
; Percent Similarity: 40.40% Conservative: 59
; Best Local Similarity: 27.37% Mismatches: 146
; Query Match: 1 Indels: 124
; DB: 1 Gaps: 19
;
; US-09-830-837-6 (1-1052) x US-08-322-677A-6 (1-1497)
;
; QY 8 LeuLeuLeuValValLeuLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
; DB 45 ATAATCTCTCTATTGGTTATTCTGCAAAATGAAAAAAGAGAGAGGATAAAGAGTGAAGGC 104
; QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
; DB 105 AAAAAGATGATGATGATGTTGCTGTTGCTTTAGCG-----TTA 143
; QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
; DB 144 ATCTTACGATGCGTTCGCGACACATCTCTCCCGCAGCGAGGAAATCAACGGG 203
; QY 53 GluTyrglyTrileValAlaPheAsnGlyTyrrPheThrAlaLysAlaAsnSerPhe 72
; DB 204 GAAAGAAATATATGTGCGGTT-----AACACACATGAGCAGC 245
; QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
; DB 246 ATGAGCGCGCTGAAGAAAGATGTCATT----- 275
; QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnLysGluLysGlnLysAla 112
; DB 276 -----TCTGAAAAAGGC 287
; QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
; DB 288 GGGAAAGTGCAAAAGCAATTCAAATATGTAGACGACGCTTCAGTCACATTAACGAAAAA 347
; QY 132 ValPheArgSerLeuLysTyrrAlaGluSerAspProThrValProCysAsnGluThrArg 151
; DB 348 GCTGTAAAAGAAATTGAAA-----AAGACCCGAGCGCTCGCT----- 383
; QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySer 171
; DB 383 ----- 383
; QY 172 GlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArg 191
; DB 384 -----TACGTTGAAGAGATCAGTACGATCGGTCAGCGCGCTCGCTTCAC 434
; QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAlaAsn 211
; DB 435 GCGGTATACAAA---ATTAAAGCCCTGCTCTGCTCCTCAAGGCTACATCGGATCAAT 491
; QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
; DB 492 GTTAAAGTAGCGTTATCGACAGCGGTATCGATCTTCTCATCCCTGCTTAAAGGTAGCA 551
; QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245

```


RESULT 7

03-06-046-753-0
; Sequence 6, Application US/08848793

; GENERAL INFORMATION:

APPLICANT: Burns, Michael E.

APPLICANT: Getty, Edward E.

; APPLICANT: Willey, Alan D.

APPLICANT: Barnett, Bobby L.

; TITLE OF INVENTION: Bleaching Compositions and Methods of Bleaching Cellulosic Materials

; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:

ADDRESSEE: The Procter & Gamb.
STREET: 11810 East River Road

STREET: 11810 East River Road
CITY: Cincinnati

; CINCINNATI

383 ----- 383
 172 GlyPheTrpHisAlaThrGlyArgHisSerArgArgLeuLeuArgAlaIleProArg 191
 384 -----TACGTGAAGAGATCAGTACGACATCGGTAGCGGAGTCCGTGCTTAC 434
 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAlaAsn 211
 435 GCGGTATCACAA---ATTAAAGCCCTGCTCTGCTCTCAAGGCTACATGATCAAT 491
 212 ValArgValAlaValPheAspThrGlyLeuSerGlyHisProHisPheLys ----- 229
 492 GTTAAAGTAGCGGTATCGACAGCGGTATCTCTCATCTGATTTAAAGGTAGCA 551
 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
 552 AGCGGAGCGCATGCTCTCTGAAACAAAT-----CCTTCCCAAGAC 596
 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 597 AACAACTCTCAGGAACCTCAGCTTGGCGGACAGTTGGGCTCTTAATACTCAATCGGT 656
 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
 657 GTATTAGCGGTGCGCCAGCCATCATCTTACGCTGTAAGTCTCGTGTGACGGT 716
 283 ValSerThrThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp 302
 717 TCGGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGATCGCAACAATATGGAC 776
 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
 777 GTATTAACTAGAGCTCGCGGACCTTCTGGTCTGCTGCTTTAAAGCGCGATGAT 836
 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 837 AAAGCC-----GTTGATCGCGGCTGATGCTGTGGCGGCGGATACGAAGCG 887
 341 Pro-----LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
 888 ACTTCGCGACGCTCAAGCACAGTGGGCTACCTCCCTGGTAAATACCTCTGTCATTCAGTA 947
 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrp 378
 948 GCGCTGTGACAGCACAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
 379 GluLeuProGlyGlyTrpGlyArgMetLysProAspIleValThrTrpGlyAlaGlyVal 398
 999 GAGCTT-----GATGTCATGGCAGCTGCGGTATCTATC 1031
 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrServalAla 416
 1032 CAAAGCACGCTTCTCGAAACAAATACGG-----GCGTACAGGCTACGTCAATGGCA 1085
 417 SerProValValAlaGlyAlaValThrLeuValSer 429
 1086 TCTCCGCGCTTGGCGGCGCTCTTGTATCTTCT 1124

RESULT 8
 US-09-445-270-1
 ; Sequence 1, Application US/09445270
 ; Patent No. 6369011
 ; GENERAL INFORMATION:
 ; APPLICANT: Rai, Saroj
 ; APPLICANT: Correa, Paul
 ; APPLICANT: Zhu, Yong
 ; APPLICANT: Graycar, Thomas
 ; APPLICANT: Bott, Richard
 ; TITLE OF INVENTION: Protease Enzymes for Tough Cleaning and
 ; COMPOSITIONS INCORPORATING SAME
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: The Procter & Gamble Company
 STREET: 11810 East Miami River Road
 CITY: Ross
 STATE: OH
 COUNTRY: USA
 ZIP: 45061
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/445,270
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Bolam, Brian M.
 REGISTRATION NUMBER: 37,513
 REFERENCE/DOCKET NUMBER: 1234
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 513/627-2457
 TELEFAX: 513/627-0318
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1497 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 US-09-445-270-1

Alignment Scores:
 Pred. No.: 8,39e-23 Length: 1497
 Score: 331.00 Matches: 124
 Percent Similarity: 40.40% Conservative: 59
 Best Local Similarity: 27.37% Mismatches: 146
 Query Match: 5.89% Indels: 124
 DB: 4 Gaps: 19

US-09-830-837-6 (1-1052) x US-09-445-270-1 (1-1497)
 QY 8 LeuLeuLeuValValLeuLeuLeuCysGlyLysHisLeuGlyAspArgLeuLys 27
 DB 45 ATAACTCTCTATTGTTATCTGCAAAATGAAAAAGGAGGATAAAGTGTAGAGGC 104
 QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
 DB 105 AAAAAAGTATGATGATGCTGCTGTTGCTTTAGCG-----TTA 143
 QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
 DB 144 ATCTTTACGATGCGGTTTCGGCAGCATCTCTGCCAGCGGAGGAAATCAACGGG 203
 QY 53 GluTrpGluTrpIleValAlaPheAsnGlyTrpPheThrAlaLysAlaArgAsnSerPhe 72
 DB 204 GAAAGAATATATTTGCGGTTT-----AACACAGACATGAGCAGC 245
 QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
 DB 246 ATGACGCGCGCTAGAGAAGATGTCATT-----TCTGAAAAAGGC 287
 QY 93 ProSerSerAspTrpProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
 DB 276 ----- 131
 QY 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
 DB 288 GGAAGAGTGAAGAGCAATTCAAATATGTAGACGCGAGCTTCATGACATTAACCAANA 347
 QY 132 ValPheArgSerLeuLysTrpValGluSerAspProThrValProCysAsnGluThrArg 151
 DB 348 GCTGTAAAGAAATTGAAA-----AAAGACCCGAGCGTCGCT----- 383

```
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 -----
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCAGTACGATCGCGGAGTCCGCGCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAGCCCTGCTCTGCACCTCTCAAGCTACACTGATCAAT 491
QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAAGTAGCGTTTCGACACGCGTATCGATCTCTCATCTGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGGAGCCAGATGGTCTCTGAACAAAT-----CCTTTCCAAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTCAGGAACTCAGTTCGGCGACAGTTGGGCTCTTAATACTCAATCGGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGGGCTTCGGCAAGCGCATCTTACGCTGTAAAGTCTCTCGGTGCTGACGGT 716
QY 283 ValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCGGGCAATACAGCTGGATCATTAACGGAATCAGTGGCGGATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAAATAGACCTCGCGGACCTCTGGTCTGCTGCTTTAAAGCGCGCAGTGTAT 836
QY 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTCATCCGGCTGCTAGTCTGTTGGCGACGCGGTAACGAAGGC 887
QY 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCAGCTCAGACAGTGGGCTACCCCTGGTAAATACCCCTTCTGTCTATTCAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GGGCTGTTCAGACAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
QY 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGGCACCTGGCGGTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGTCTCTGGAAACAATAACGGG-----GCTACACAGGTAGTCAATGGCA 1085
QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGACGCTTCGGGAGCGGCTGCTTGTGATCTTCT 1124
```

RESULT 9

```
US-09-178-155-1
; Sequence 1, Application US/09178155
; Patent No. 6312936
; GENERAL INFORMATION:
; APPLICANT: Poulos, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nachtern, Joanne
; APPLICANT: Nakl, Donald P.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2
```

```
; CURRENT APPLICATION NUMBER: US/09/178,155
; CURRENT FILING DATE: 1998-10-23
; EARLIER APPLICATION NUMBER: 08/956,323
; EARLIER FILING DATE: 1997-10-23
; EARLIER APPLICATION NUMBER: 08/956,564
; EARLIER FILING DATE: 1997-10-23
; EARLIER APPLICATION NUMBER: 08/956,324
; EARLIER FILING DATE: 1997-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
```

```
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B. amyloliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
US-09-178-155-1
```

Alignment Scores:

```
Pred. No.: 8,54e-23 Length: 1513
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: 4 Gaps: 19
```

US-09-830-837-6 (1-1052) x US-09-178-155-1 (1-1513)

```
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGTCTATTGGTTATCTGCAATGAAAAAGGAGAGGATAAAGAGTGAGAGGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGTATGATCATCTTGTCTTGTCTTACGC-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTACGATGGCTTCGGCAGCAGCATCTCTCCAGCGGCGGAGGAAATCAACGGG 203
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATATATGCGGGTT-----AAACAGACAATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIlelleProArgAsn 92
Db 246 ATGAGCGCGCTAAGAAAGAAAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287
QY 113 GlyLeuLeuThrLeuGlu-----AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAAGTGCAGAAAGCAATATCAAAATATGTAGAGCGAGCTTCAGCATTAACGAAAAA 347
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAAGAAATTGAAA-----AAAGACCGGAGCGTCT 383
QY 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 ----- 383
QY 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCAGTACGATCGCGGAGTCCGCGCTTAC 434
QY 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATACAA--ATTAAGCCCTGCTCTGCTCTCTCAAGGCTACACTGATCAAT 491
```

QY 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAAGTAGGGTTATCGACAGCGGTATCATCTTCATCTCGATTTAAAGGTAGCA 551
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGAGCAGCATGGTCTCTTCTGAACAAT-----CCTTTCCAGAC 596
QY 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTACGGAACATCACGTTCCGGCACAGTTGCGCTCTTAATAACTCAATCGT 656
QY 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGCGTTGCGCAAGCGCATCTTACGCTGTAAGTTCTCGGTGCTGACGCT 716
QY 283 ValSerTrpSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp 302
Db 717 TCGGCAATACAGCTGGATCAATTAAGGATCGAGTGGCGATCGCAACAATATGGAC 776
QY 303 ValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTACATGAGCTCTCGCGACCTTCTGCTTCTGCTTAAAGCGGAGTTGAT 836
QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAGCC-----GTTGATCTCGGCTGTAGTGTTCGGGACGCGGTACGAGGC 887
QY 341 Pro-----LeuTrpGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCAGCTCAAGCACAGTGGCTGACCTGGTAAATACCTCTGTCATTCAGTA 947
QY 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrp 378
Db 948 GCGCTGTGTGACGACGACCAACCAAGACATCTTCTCAAGCGTAGGACCT----- 998
QY 379 GluLeuProGlyGlyTrpGlyArgMetLysProAspIleValThrTrpGlyValVal 398
Db 999 GAGCTT-----CATGTCATGGCAGCTGGCTATCTATC 1031
QY 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGCTCTCTGGAACAATAACGCG-----CGGTACAAGGTACGTCAATGGCA 1085
QY 417 SerProValAlaGlyAlaValThrLeuValSer 429
Db 1086 TCTCGCAGCTTCCGGAGCGGCTGCTTGTATCTTCT 1124

RESULT 10

US-09-000-016-1

Sequence 1, Application US/09000016

Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.

TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV

TITLE OF INVENTION: ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/000,016

FILING DATE: January 30, 1998

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1

Alignment Scores:

Pred. No.: 3,14e-22 Length: 2809
Score: 330.00 Matches: 169
Percent Similarity: 36.98% Conservative: 81
Best Local Similarity: 25.00% Mismatches: 258
Query Match: 5.88% Indels: 169
DB: 3 Gaps: 28

US-09-830-837-6 (1-1052) x US-09-000-016-1 (1-2809)

QY 125 ArgValThrProGlnArgLysValPheArgSerLeuLysTyrrAlaGluSer----- 141
Db 745 CGTCGGCTACCGGGCCGCCAGCGGCCGCCCAAGCGCGGTCGCGAAGCGGGGAACT 804
QY 142 -----AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSer 159
Db 805 CCGCGGACCTCTGACGTCCT-----GAACCGGACGCGGTCCG 843
QY 160 ArgProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheThrHisAlaThrGlyArg 179
Db 844 GACCCCGCACGA-----GGACGCGTCCGAGCTGTGGGACGC 879
QY 180 HisSerSerArgArgLeuLeuArg----- 187
Db 880 CGTCACCAACCGGACCGGACCGGATCGCGGATCGCGGATCGCGGATCGCGGATCGCG 939
QY 188 ---AlaIleProArgGlnValAlaGlnThr-LeuGlnAlaAspValLeuTrpGlnMetG1 206
Db 940 CAGGCGCCCTCGACACGTCGCGGCGAGATCGCGGCGGATCGCGGCGGATCGCGGCGG 999
QY 206 YTrpThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPr 226
Db 1000 CTACGACGCGCAAGGGGTGAGATCGCGTCTGACACCGGTGTCGACACGAGCATCC 1059
QY 226 ohisPhelys---AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAs 245
Db 1060 GCACCTGAGGCGGCGGTGACCGGTCCCAAGATTCACCGCGCGCGCGCGCGCGCGA 1119

Db 1194 TCCTTCTACTATGGAAGGCTGTGATCAACGTACAGCGCGCAGCTCAG 1241

RESULT 14

5472855-1

Patent No. 5472855

APPLICANT: CARTER, PAUL J.; WELLS, JAMES A.

TITLE OF INVENTION: SUBSTRATE ASSISTED CATALYSIS

NUMBER OF SEQUENCES: 31

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/287,964

FILING DATE: 22-SEP-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 90,902

FILING DATE: 12-JUL-1993

APPLICATION NUMBER: 823,039

FILING DATE: 14-JAN-1992

APPLICATION NUMBER: 35,652

FILING DATE: 06-APR-1987

APPLICATION NUMBER: 334,081

FILING DATE: 04-APR-1989

APPLICATION NUMBER: 127,134

FILING DATE: 01-DEC-1987

APPLICATION NUMBER: 846,627

FILING DATE: 01-APR-1986

APPLICATION NUMBER: 858,594

FILING DATE: 30-APR-1986

APPLICATION NUMBER: 614,612

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,615

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,617

FILING DATE: 29-MAY-1984

APPLICATION NUMBER: 614,491

FILING DATE: 29-MAY-1984

SEQ ID NO:1:

LENGTH: 1497

5472855-1

Alignment Scores:

Pred. No.:	1,48e-22	Length:	1497
Score:	328.50	Matches:	130
Percent Similarity:	40.32%	Conservative:	70
Best Local Similarity:	26.21%	Mismatches:	169
Query Match:	5.85%	Indels:	127
DB:	6	Gaps:	21

US-09-830-837-6 (1-1052) x 5472855-1 (1-1497)

QY	8	LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys	27
Db	45	ATAATCTGTCTATTGGTTATCTGCAAAATGAAAAAGGAGAGGATAAAGAGTGAGAGGC	104
QY	28	LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu	41
Db	105	AAAAAGTATGATCATCTGCTGTTGCTTTAGCG-----TTA	143
QY	42	ThrLeuLysValGluPheSerSerThrValVal-----	52
Db	144	ATCTTTACGATGGCGTTGCGGCACACATCTCTGCGCGCGCGGAAATCAACGGG	203
QY	53	GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe	72
Db	204	GAAGAAGATATATTCGCGGTTT-----AAACAGACAATGAGCAGC	245
QY	73	IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn	92
Db	246	ATGAGCGCGCTAAGAAGAAAGATGTCATT-----TCTGAA	281
QY	93	ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla	112
Db	282	AAAGCGGGAAGTGAAGCAATTCAAATATGATAGC-----GCA	323
QY	113	GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal	132

Db	324	GCTTCAGCTACATTAAACGAA-----	344
QY	133	PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp	152
Db	345	-----AAAGCTGTAAGAAATTTGAAAAAGACCCGAGCGTCGCT-----	383
QY	153	SerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGly	172
Db	383	-----	383
QY	173	PheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArgGln	192
Db	384	-----TACGTTGAAGAAGATCAGTAGCACATCGGTACGCGAGTCGCTTACGGC	437
QY	193	ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal	212
Db	438	GTATCAAA---ATTAAAGCCCTGCTCTGCATCTCTCAAGCTACATCGATCAATGTT	494
QY	213	ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys-----	229
Db	495	AAAGTAGCGGTTATCGACACGCGTATCGATTCTTCATCCTGATTTAAAGGTAGCAGC	554
QY	230	-----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly	246
Db	555	GGAGCCAGCATGTTCTCTGAAACAAAT-----CCTTTCCAAGACAAC	599
QY	247	LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys	263
Db	600	AACTCTCAGGAACCTCAGTTCGCGCACAGTTCGCGCTCTTAATAACTCAATCGGTGTA	659
QY	264	GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal	283
Db	660	TTAGCGGTTGCGCAACGCGATCATTACGCTGTAAAGTTCTCGTGTGCGGTTC	719
QY	284	SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal	303
Db	720	GGCCAATACAGCTGGATCATTAACGGAATCAGTGGCGCATCGCAACAATATGGAGTT	779
QY	304	LeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAspLys	321
Db	780	ATTAACATGAGCTCGCGGAGCCTCTGCTGCTGCTGCTTAAAGCGCGCATGTATATA	839
QY	322	ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro	341
Db	840	GCC-----GTGATCCGCGCTGCTAGTGTGCGCGCAGCGCTAAGAGGCACT	890
QY	342	-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly	359
Db	891	TCCGGCAGCTCAAGCACAGTGGGTACCTGGTAAATACCCCTTCTGTCATTGCGTAGGC	950
QY	360	GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu	379
Db	951	GCTGTTGACAGCAGCAACAAAGAGCATCTTCTCAAGCTAGGACCT-----GAG	1001
QY	380	LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg	399
Db	1002	CTT-----GATGTCATGCGCACCTGCGGTATCTATCCAA	1034
QY	400	GlySer-----GlyValLysGlyLysArgAlaLeuSerGlyThrSerValAlaSer	417
Db	1035	AGCACGCTCTCTGAAACAAATACGGG-----GCGTACACACGCTAGCTCAATGGCATCT	1088
QY	418	ProValValAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal	437
Db	1089	CCGACGTTGCCGAGCGGCTGCTTGTATCTTTCT-----AAGCACCCGAACTGG	1139
QY	438	AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn	457
Db	1140	ACAAACACTCAAGTCCGACGAGCTTTAGAAAACACCACCTACAAACTTGGT-----GAT	1193
QY	458	MetPheGlnGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln	473

Db 1194 TCTTTTACTATGAAAGGGGTGATCAACGTACAGCGGCGAGCTCAG 1241

RESULT 15

US-08-069-863-1
Sequence 1, Application US/08069863
Patent No. 5470733
GENERAL INFORMATION:
APPLICANT: BRYAN, Philip N
APPLICANT: ALEXANDER, Patrick
APPLICANT: STRAUSSBERG, Susan L
TITLE OF INVENTION: CALCIUM FREE SUBTILISIN MUTANTS
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/069.863
FILING DATE: 01-JUN-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Meuth, Donna M
REGISTRATION NUMBER: 36,607
REFERENCE/DOCKET NUMBER: 028755-016
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 450..1599
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 772..1599
FEATURE:
NAME/KEY: misc_feature
LOCATION: 450
OTHER INFORMATION: /note= "Amino Acid Val at position 450 is fMet."
US-08-069-863-1

Alignment Scores:
Pred. No.: 2,18e-22 Length: 1868
Score: 328.50 Matches: 130
Percent Similarity: 40.32% Conservative: 70
Best Local Similarity: 26.21% Mismatches: 169
Query Match: 5.85% Indels: 127
Gaps: 21

US-09-830-837-6 (1-1052) x US-08-069-863-1 (1-1868)

Qy 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGlyLys 27
Db 399 ATAAATCTGTCTATTGTTATTCGAAATGAAAAAGAGAGGATTAAGAGTGAGAGC 458
Qy 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 459 AAAAAAGTATGGATCATGTTGCTGTTGCTTTAGCG-----TTA 497

Qy 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 498 ATCTTTTACGATGGCTTCGCGACACATCTCTCCAGCGCGAGGAAATCAACGGG 557
Qy 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaAsnSerPhe 72
Db 558 GAAAGAAATATATTCGCGTT-----AAACAGACAATGAGCAGC 599
Qy 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
Db 600 ATGACGCCCGCTTAAGAGAAAGATGTCATT-----TCGAA 635
Qy 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 636 AAGCGGGGAAAGTGCAAAAGCAATTCAAATATGTAGC-----GCA 677
Qy 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
Db 678 GCTTCAGCTACATTAACGAA----- 698
Qy 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
Db 699 ---AAAGCTGTAAAGAAATTTGAAAAGACCCGAGCTCGCT----- 737
Qy 153 SerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
Db 737 ----- 737
Qy 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 738 -----TACGTTGAAGAAGATCACGTAGCACATCGTACGCGCAGTCGCTTACGGC 791
Qy 193 ValAlaGlnThrLeuGlnAlaAspValLeuLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 792 GTATCACAA---ATTAAAGCCCTCTCTGCACCTCTCAAGGCTACATGGATCAATGTT 848
Qy 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 849 AAGTAGCGGTTATCGACAGCGGATCGATCTTCCTCCTCCTCCTGATTTAAAGTAGCAGGC 908
Qy 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db 909 GGAGCCAGCATGGTCTCTCTGAAACAAAT-----CCTTCCAGAGCAAC 953
Qy 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 954 AACTCTACGGAACACTCAGTTGCGGCGACAGTTGCGGCTCTTAATAACTCAATCGGTGTA 1013
Qy 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 1014 TTAGCGGTTGCGCAAGCGCATCACTTTACGCTGTTAAAGTTCTCGGTGCTGAGCGGTCC 1073
Qy 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 1074 GGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGCATCGCAACAATATGACGTT 1133
Qy 304 LeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db 1134 ATTAACATGAGCTCGCGGACCTTCTGGTTCTGCTGCTTTAAAGCGGCGAGTTGATAAA 1193
Qy 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
Db 1194 GCC-----GTTGCATCCGCGCTGCTGCTGCGGCGCGCCGTAACGAGGCACT 1244
Qy 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 1245 TCCGGCAGCTCAAGACACATGGCTGACCTGGTAAATPACCTTCTGTCTATTCGATGAGGC 1304
Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 1305 GCTGTTGACAGCAGCAACCAAGAGCATCTTTCTCAAGCTAGGACCT-----GAG 1355
Qy 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrThrGlyAlaGlyValArg 399

```
Db 1356 CTT-----|||:||||:|GATGTCATGCACCTGGCGTATCTATCCAA 1388
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db 1389 AGCAGCTTCCTGGAAACAATACGG-----GCGTACAACGGTACGTCATATGGCATCT 1442
QY 418 ProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db 1443 CGCGAGCTTCCGGAGCGGCTGTTTGATCTCTTCT-----AAGCACCCGAACTGG 1493
QY 438 AsnProAlaSerMetLysGlnAlaLeuLeuAlaSerAlaArgLeuProGlyValAsn 457
Db 1494 ACAACACTCAAGTCGCGAGCGAGTGTAGAAAACACCACTACAAAACCTGGT-----GAT 1547
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db 1548 TCTTCTACTATGGAAAGGGGTGATCAACGTACAGCGCGGCGACCTCAG 1595

RESULT 16
US-08-309-069-1
; Sequence 1, Application US/08309069
; Patent No. 5567601
; GENERAL INFORMATION:
; APPLICANT: BRYAN, Philip N
; APPLICANT: ALEXANDER, Patrick
; APPLICANT: STRAUSBERG, Susan L
; TITLE OF INVENTION: CALCIUM FREE SUBTILISIN MUTANTS
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/309,069
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028755-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note= "Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
US-08-309-069-1
Alignment Scores:
```

```
Pred. No.: 2,18e-22 Length: 1868
Score: 328.50 Matches: 130
Percent Similarity: 40.32% Conservative: 70
Best Local Similarity: 26.21% Mismatches: 169
Query Match: 5.85% Indels: 127
DB: Gaps: 21

US-09-830-837-6 (1-1052) x US-08-309-069-1 (1-1868)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db 399 ATAATCTGCTATTGTTATTCGCAATGAAGAAAGGAGAGAGATAAGAGTGAGAGGC 458
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 459 AAAAAGATGATGATCATGTTTGTGTTTCTTTAGCG-----TTA 497
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 498 ATCTTTACGATGGCGTTTCGGCAGCACATCTCTGCCAGCGCAGGAAATCAACGGG 557
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 558 GAAAGAAATATATTCTCGGTTT-----AAACAGACAATGAGCAGC 599
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTyrArgIleIleProArgAsnAsn 92
Db 600 ATGAGCGCGCTAAGAAAGATGTCATT-----TCGAA 635
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 636 AAAGCGGGAAAGTCAAAAGCAATCAAAATATGTAGAC-----GCA 677
QY 113 GlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
Db 678 GCTTCAGTACATTAAACGAA----- 698
QY 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTyr 152
Db 699 ---AAAGCTGTAAGCAATGAAAGAACCCGAGGTCGCT----- 737
QY 153 SerGlnLysTyrGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
Db 737 ----- 737
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 738 -----TACGTTGAAGAAGATCACGTAGCACATGCGTACGCGCAGTCCGCTTACGGC 791
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTyrGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 792 GTATCACA---ATTAAAGCCCTGCTCTGCACCTCTCAAGGCTACACTGGATCAATGTT 848
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 849 AAAGTAGCGGTTATCGACAGCGGTATCGATTCTTCTCTCATCTCTGATTAAAGGTAGCAGC 908
QY 230 -----AsnValLysGluArgThrAsnTyrThrAsnGluArgThrLeuAspAspGly 246
Db 909 GGAGCCAGCATGTTCTCTCTGAAACAAT-----CCTTTCAAGACAAC 953
QY 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 954 AACTCTCAGGAACTCACGTTCCGGCAGAGTTCGGCTTAAATCAATCAATCGGTGTA 1013
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 1014 TTAGCGGTTCCGCCAAGCGCATCACATTTACGCTGTAAAGATTCTCGTGTGCGGTTCC 1073
QY 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 1074 GGCCAATACAGCTGGATCATTAACGGAATCGAGTGGCGCATCGCAAAACAATATGGACGT 1133
```



```
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
849 AAAGTAGCGTATCCACAGCGTATCGATTCTCTCATCCTGATTAAAGGTAGCAGGC 908
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
909 GGAGCCAGCAGTGTCTCTCTCAACAAT-----CCTTTCCAAGACAAC 953
QY 247 LeuGlyHisGlyThrPheValAlaGlyValLleAlaSerMetArgGlu-----Cys 263
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
954 AACTCTCAGGAAGTACAGTTCGCGCAGATTCGCGCTCTTAATCACTCAATCGGTGA 1013
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisLlePheArgValPheThrAsnAsnGlnVal 283
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1014 TTAGCGCTTCGCGCAAGCAGTACATTTACGCTGTAAAGTTCTCGTGTGCGGTTC 1073
QY 284 SerTyThrSerTrpPheLeuAspAlaPheAsnTyAlaLleLeuLysLysLysVal 303
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1074 GGCAATACAGCTGGATGATCAATTACGAATCGAGTGGCGATCCAAACAATATGACGT 1133
QY 304 LeuAsnLeuSerLleGlyGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1134 ATTAACATGAGCTCGCGGACCTCTGTTCTGCTGCTTTAAAGCGCGAGTTGATAA 1193
QY 322 ValTrpGluLeuThrAlaAsnAsnValLleMetValSerAlaLleGlyAsnAspGlyPro 341
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1194 GCC-----GTTGCATCCGCGCTGCTAGTGTGCGCAGCCGTAACGAAGCACT 1244
QY 342 -----LeuTyGlyThrLeuAsnAsnProAlaAspGlnMetAspValLleGlyValGly 359
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1245 TCCGCGAGCTCAAGCAGACAGTGGCTACCTGCTGTAATACCTTCTCATTCAGTAGGC 1304
QY 360 GlyIleAspPheGluAsnLleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1305 GCTGTTGACAGCAGCAACAGAGCATTTCTCAAGCTAGGACCT-----GAG 1355
QY 380 LeuProGlyGlyTyGlyArgMetLysProAspLleValThrTyGlyAlaGlyValArg 399
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1356 CTT-----GATGTCATGCGACCTGGCGTATCTATCCAA 1388
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1389 AGCAGCTCTCTGGAACAATACGGG-----GCGTACACGGTACGTCATGGCATCT 1442
QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1443 CCGCAGTTCGCGGAGCGCTGCTTGTATCTTCT-----AGCACCCGACTGG 1493
QY 438 AsnProAlaSerMetLysGlnAlaLeuLleAlaSerAlaArgArgLeuProGlyValAsn 457
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1494 ACAACACTCAAGTCCGCGCAGCTTTAGAAAACACCACCTACAAAACCTTGGT-----GAT 1547
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrgln 473
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
1548 TCTTCTACTATGAAAGGCGTGTATCAAGTACAGCGGCGACTCAG 1595
```

RESULT 18

PCT-US95-05520-1

; Sequence 1. Application PC/TUS9505520

; GENERAL INFORMATION:

; APPLICANT: BRYAN, Phillip N

; APPLICANT: ALEXANDER, Patrick

; APPLICANT: STRAUSBERG, Susan L

; TITLE OF INVENTION: CALCIUM FREE SUBTILISIN MUTANTS

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Burns, Doane, Swecker & Mathis

; STREET: P.O. Box 1404

; CITY: Alexandria

; STATE: Virginia

; COUNTRY: United States

; ZIP: 22313-1404

; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05520
; FILING DATE: 28-APR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/309,069
; APPLICATION NUMBER:
; FILING DATE: 20-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meuth, Donna M
; REGISTRATION NUMBER: 36,607
; REFERENCE/DOCKET NUMBER: 028758-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-8620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1868 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 450..1599
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 772..1599
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 450
; OTHER INFORMATION: /note= "Amino Acid Val at position
; OTHER INFORMATION: 450 is fMet."
; PCT-US95-05520-1

Alignment Scores:
Pred. No.: 2,18e-22 Length: 1868
Score: 328.50 Matches: 130
Percent Similarity: 40.32% Conservative: 70
Best Local Similarity: 26.21% Mismatches: 169
Query Match: 5.85% Indels: 127
DB: 5 Gaps: 21

US-09-830-837-6 (1-1052) x PCT-US95-05520-1 (1-1868)

QY 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
399 ATAATCTGCTATTGTTATTCTGCAAAATCAAAAAAGGAGAGGATAAAGAGTCAGAGGC 458
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
459 AAAAAAGTATGGATCATGTTTGTCTTGTAGCG-----TTA 497
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
498 ATCTTTACGATGCGGTTCGCACACATCTCTGCCCGCGCAGGGAATCAACGGG 557
QY 53 GluTyGluTyTrileValAlaPheAsnGlyTyPheThrAlaLysAlaArgAsnSerPhe 72
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
558 GAAAGAAATATATTTCGGGGTT-----AAACAGACAATGAGCAGC 599
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIlelleProArgAsnAsn 92
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
600 ATGAGCGCGCTAAGAAAGAAAGATGTCATT-----TCTGAA 635
QY 93 ProSerSerAspTyProSerAspPheGluValLleGlnLleLysGluLysGlnLysAla 112
Db :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
636 AAAGCGGGGAAGTGCAGAAAGCAATCAATATGTAGAC-----GCA 677
QY 113 GlyLeuLeuThrLeuGluAspHisProAsnLleLysArgValThrProGlnArgLysVal 132
```



```
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 151 ATCTTTACGATGCGTTCGCCACACATCTCTGCCCGCCGCGGAAATCAACGGG 210
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 211 GAAAGAAATATATCTCGGGTTT-----AAACAGACAATGAGCAGC 252
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
Db 253 ARGAGCCCGCTAAGAAGAAAGATGCAAT-----TCGAA 288
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 289 AAAGCGCGAAAGTGAAGCAATTCAAATATGTAGAC-----GCA 330
QY 113 GlyLeuLeuThrLeuLysPheProAsnIleLysArgValThrProGlnArgLysVal 132
Db 331 GCTTCAGCTACATTAACGAA----- 351
QY 133 PheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
Db 352 ---AAAGCTGTAAAGAAATGAAAGACCGGAGCTGCT----- 390
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySerGly 172
Db 390 ----- 390
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGln 192
Db 391 -----TAGCTTGAAGAAGATCAGCAGTACATCGCTAGCGGAGTCCGCTACGGC 444
QY 193 ValAlaGluThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVal 212
Db 445 GTATCAAA---ATTAAAGCCCTGCTCTGCACTCTCAAGGCTACACTGGATCAATGTT 501
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 502 AAAGTAGCGGTATCGACAGCGTATCGATTCTCTCATCTCTTAAGGTAGCAGC 561
QY 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGly 246
Db 562 GGAGCCAGCATGTTCTCTTCAACAAAT-----CCTTTCCAAGACAAC 606
QY 247 LeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263
Db 607 GACTCTCAGGAACTACGTTCCGCCACAGTTGCGGCTCTTAATCACTCAATCGGTGTA 666
QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnVal 283
Db 667 TTAGCGGTGCGCCAAAGCATCATTACGCTGTAAAGTTCGCTGCTGACGGTTC 726
QY 284 SerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAspVal 303
Db 727 GGCCAAATACAGTGTGATCATTAACGAATCGAGTGGCGCATGCAACAATATGGAGTT 786
QY 304 LeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheValAspLys 321
Db 787 ATTAACATGAGCTCGCGGACCTTCTGTTCTGCTGTTTAAAGCGGCGATGTATAAA 846
QY 322 ValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
Db 847 GCC-----GTTGATCCGCGCTGCTAGTGTGCTGCGGACCGGTAACGAGCAT 897
QY 342 -----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 898 TCCGCGAGCTCGTCACAGCTGACTACCTGCAAAATACCTCTCTCATTCAGTATGAGC 957
QY 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 958 GCTGTTGACAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT-----GAG 1008
```

```
QY 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg 399
Db 1009 CTT-----GATGTATGCGCACTGCGGTATCTATCAAA 1041
QY 400 GlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db 1042 AGCAGCTCTCTCGAAACAATAACGGG-----CGGTACAACGGTACCTCAATGGCATCT 1095
QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db 1096 CCGACGTTGCCGAGCGGCTGCTTGATTTCTTCT-----AAGCACCGCACTGG 1146
QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
Db 1147 ACAACACACTCAAGTCCGCGAGCAGTTTAGAAAACACCACTACAAAACCTTGT-----GAT 1200
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db 1201 TCTTTCTACTATGAAAGGGCTGATCAACGTACAGCGGCGCTCAG 1248

RESULT 22
RE34606-4
; Patent No. RE34,606
; APPLICANT: ESTELL, DAVID A.; WELLS, JAMES A.; BOTT,
; RICHARD R.
; TITLE OF INVENTION: MODIFIED ENZYMES AND METHODS FOR
; MAKING SAME
; NUMBER OF SEQUENCES: 27
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/556,918
; FILING DATE: 20-JUL-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 614,612
; FILING DATE: 29-MAY-1984
; SEQ ID NO: 4:
; LENGTH: 1498
RE34606-4

Alignment Scores:
Pred. No.: 2,94e-22 Length: 1498
Score: 325.50 Matches: 99
Percent Similarity: 47.58% Conservative: 58
Best Local Similarity: 30.00% Mismatches: 128
Query Match: 5.79% Indels: 45
DB: 6 Gaps: 14

US-09-830-837-6 (1-1052) x RE34606-4 (1-1498)

QY 180 HisSerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAla 199
Db 438 CATATTGCACATGAATATGCGCAATCTGTCCTTATGGCATTTCTCAA---ATTAAAGCG 494
QY 200 AspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThr 219
Db 495 CCGGCTCTTCACTCTCAAGGCTACACAGGCTTAACGTAAAGTAGCTGTATCGCAGC 554
QY 220 GlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsn 239
Db 555 GGAATGACTCTCTCTCTGACTTA---AAGCTCAGAGCGGCGGACGACATTCGACCT 611
QY 240 GluArgThr-----LeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIle 257
Db 612 TCTGAAACAACCCATACCAGGACGCGCAGTCTTCACGGTACGATGTAGCCGTACGATT 671
QY 258 AlaSerMetArgGlu-----CysGlnGlyPheAlaProAspAlaGluLeuHisIle 274
Db 672 GCGGCTCTTAATAACTCAATCGGTGTTCTGGGGTGTAGCCCAAGCGCATCATATATGCA 731
QY 275 PheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsn 294
Db 732 GTAAAGTCTGATTCAACAGAGCGGCCAATATAGCTGGATTATTACGGCATTTAG 791
QY 295 TyrAlaIleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyPro-----Asp 312
```



```
QY 233 uArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheVa 253
Db 465 CGAGACCAAC-----CCGTTCCAGGACACCAACAGCCATGGCCACCCACGT 509
QY 253 lAlaGlyValIleAlaSerMetArgGlu-----CysGlnGlyPheAlaProAspAl 270
Db 510 CCGCGGACCGTGGCGCGCTCCACCAACAGCATCGCGCTCGTGGCGTGGCGCGCGAGCG 569
QY 270 aGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLe 290
Db 570 CAGCCTCTAGCGGTCAAGTACTCGCGCGCCAGCGGCGGCGGCGGCGGCGGCGGCGGAT 629
QY 290 uAspAlaPheAsnTyrAlaIleLeuLysLysIleAspValLeuAsnLeuSerIleGlyG 310
Db 630 CAAGGCGCATGAGTGGGCGCATGCCAACACATGGAGCGTCATCACCATGAGCTCGCGG 689
QY 310 yProAspPheMetAsp-----HisProPheValAspLysValTrpGluLeuThrAlaAs 328
Db 690 CCGAGCGGCGAGCGCGCGCTCAAGCGCGCGCTCGACAAAGGCC-----GTCGCCAG 740
QY 328 nasnValIleMetValSerAlaIleGlyAsnAspGlyPro-----LeuTyrGlyThrLe 346
Db 741 CGGCGTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 800
QY 346 uAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIleAspPheGluAspAs 366
Db 801 CGGCTACCGCGGCGAGTACCGAGCGTCATCGCGCGCGCGCGCGCGCGCGCGCGCG 860
QY 366 nIleAlaArgPheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGlyAr 386
Db 861 GCGCGCGAGCTTCAGCGGTCGCGCGCG-----GAGCTG-----894
QY 386 gMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySer-----GlyVally 404
Db 895 -----GAGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 944
QY 404 sGlyGlyCysArgAlaLeuSerGlyThrSerValAlaLeuSerProValValAlaGlyAlaVa 424
Db 945 GTACGGC-----GCCTACAGCGGCGCACAGCATGGCGCGCGCGCGCGCGCGCGCG 998
QY 424 lThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLysG 444
Db 999 TGCATCTATCTCTCAGC-----AAGCACCGCGACCTGGACCAACACCGCGCGCG 1049
QY 444 nAlaLeuIleAlaSerAlaArgLeuProGlyValAlaAsnMetPheGluGlnGlyHisG 464
Db 1050 CAGCTTGGAGACACCAACCGACCGAGCTCGCG-----GACAGCTTCTACTACGCGAAGG 1103
```

RESULT 24

US-08-397-602A-8

; Sequence 8, Application US/08397602A

; Patent No. 5646044

; GENERAL INFORMATION:

; APPLICANT: Wilson, Charles R

; APPLICANT: Tang, Maria R

; APPLICANT: Berger, Harald

; APPLICANT: Christianson, Teresa M

; APPLICANT: Hansen, Dieter

; TITLE OF INVENTION: Expression Systems for the Production

; of Target Proteins in Bacillus

; NUMBER OF SEQUENCES: 9

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Henkel Corporation Law Department

; STREET: 140 Germantown Pike, Suite 150

; CITY: Plymouth Meeting

; STATE: PA

; COUNTRY: USA

; ZIP: 19462

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/397,602A

FILING DATE: 02-MAR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jaeschke, Wayne C

REGISTRATION NUMBER: 21,062

REFERENCE/DOCKET NUMBER: D8969/M4828

TELECOMMUNICATION INFORMATION:

TELEPHONE: (610) 832-2200

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1452 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: other nucleic acid

DESCRIPTION: /desc = "DNA sequence of the

DESCRIPTION: alkaline protease gene of strain ATCC 53926 and its controlling

DESCRIPTION: elements"

ORIGINAL SOURCE:

ORGANISM: Bacillus licheniformis

STRAIN: ATCC 53926

US-08-397-602A-8

Alignment Scores:

Pred. No.: 3,92e-22 Length: 1452

Score: 324.00 Matches: 133

Percent Similarity: 41.14% Conservative: 69

Best Local Similarity: 27.09% Mismatches: 183

Query Match: 5.77% Indels: 109

DB: 1 Gaps: 17

US-09-830-837-6 (1-1052) x US-08-397-602A-8 (1-1452)

```
QY 2 LysLeuValAsnIle---TrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
Db 159 AGAATAATTCGATAGATTATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 218
QY 21 LeuGlyAspArgLeuGlyLysSerPheGluLysAlaProCys-----ProGlyCys 38
Db 219 AGAGTCAGTAATGATGAGGAAAGAGATTTT---GGCTTGGGATGCTGACGGCTTCATGC 277
QY 39 SerHisLeuThrLeuLysValGluPheSerSerThrVal-----Val 52
Db 278 TCGTG-TTCACGATGGCATTCAGCGATTCGCGTCTCTCTCTCTCTCTCTCTCTCTCTCT 336
QY 53 GluTyrGluTyrIleValAlaPhe---AsnGlyTyrPheThrAlaLysAlaArgAsnSer 71
Db 337 GAAAGGATTATATTCGGATTAAAGTCAGGAGTGAACCCGCATCTGTCAAAAGGAC 396
QY 72 PheIleSerSerAlaLeuLysSerSerGluValAsp---AsnTrpArgIleIleProArg 90
Db 397 ATCATCAAGAGAGC-----GCCGGAAGAGTGACAGCAGTTAGATCATCAACGCG 450
QY 91 AsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGln 110
Db 451 GCAAAAGCGAAGCTAGAC-----468
QY 111 LysAlaGlyLeuLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArg 130
Db 469 AAAGAAGCGCTTAAGGAAGTCAAAATGATCCGGATGTC-----507
QY 131 LysValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThr 150
Db 508 -----GCTTATGCGAAGAGGAT-----525
QY 151 ArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGly 170
```

Db 525 ----- 525
Qy 171 SerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuArgAlaIlePro 190
Db 526 -----CATGTGGCCATCGCTTGCGGCAACCGTCTC 558
Qy 191 ArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAla 210
Db 559 TAGGGCATT---CCTCTCATTAAGCGGACAAAGTCAGGCTCAAGGCTTAAGGAGCG 615
Qy 211 AsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsn 230
Db 616 ATGTAAAGTAGCGCTCTGGATACAGGATCAAGCTTCTCATCCGCACTTGAACGTA 675
Qy 231 ValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGly 250
Db 676 GTGGCGGAGCAAGCTTTGTGGCTGGCGGAAGCTTATAACACCGACGCAAGGACGCG 735
Qy 251 ThrPheValAlaGlyValIleAlaSerMetArgGluCysGln-----GlyPheAla 267
Db 736 ACACATGTTGGCGTACAGTAGCTGCGTTGACAATAACAAGGGTGTATTAGCGTTGCG 795
Qy 268 ProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSer 287
Db 796 CCAAGCGTATCTTGTACGCGTTAAAGTACTGAATTAAGCGGAAGCGGATCATACAGC 855
Qy 288 TrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSer 307
Db 856 GGCATTGTAAGCGAATCGGTGGCGGACACAAACCGCATGCGATGATATCATATAGAC 915
Qy 308 IleGlyGlyProAspPheMetAsp-----HisProPheValAspLysValTrpGluLeu 325
Db 916 CTTGGGGAGCATCAGCGTCGACAGGATGACAGGAGCGATGACATGATAT----- 969
Qy 326 ThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGlyProLeu-----Tyr 343
Db 970 ---GCAAGAGGGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1026
Qy 344 GlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIleAspPhe 363
Db 1027 AATACAAATGGCTATCTCTGCAAAATACGATTTCTGTCATCGTGTGTGTGTGTGTGT 1086
Qy 364 GluAspAsnIleAlaArgPheSerArgGlyMetThrThrTrpGluLeuProGlyGly 383
Db 1087 AACAGCAACAGAGCTTCATTTTCCAGCGTCGGAGCA-----GAGCTT----- 1128
Qy 384 TyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyVal 403
Db 1129 -----GAAATCATGCTCTGCGGCGAGCGGTATACAGCACTTACCAC 1170
Qy 404 LysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAla 423
Db 1171 ACGAACACTTATGCAACATTGAACGGAAGCTCAATGGCTTCTCTCATGTAGCGGAGCA 1230
Qy 424 ValThrLeuLeuValSer-ThrValGlnLysArgGluLeuValAsnProAlaSerMetIly 443
Db 1231 GCAGCTTGTGATCTGTGCAAAACATCCGAACCTTTCAGCTT-----CACAA 1275
Qy 443 sGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGluGlnGlyHI 463
Db 1276 GTCCGACACCGTCTCTCCAGCAGCGGCACTTATTTGGGA-AGCTCTTCTACTATGGAA 1334
Qy 463 sGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db 1335 AGGCTGTGATCAATGTGAAGTGGCGGTCAA 1365
RESULT 25
US-09-000-016-3
; Sequence 3, Application US/09000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: AKIRA ARISAWA et al.

; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2539 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces viridosporus
; STRAIN: A-914
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; US-09-000-016-3
Alignment Scores:
Pred. No.: 1.03e-21 Length: 2539
Score: 324.00 Matches: 116
Percent Similarity: 42.46% Conservative: 50
Best Local Similarity: 29.67% Mismatches: 154
Query Match: 5.77% Indels: 71
DB: 3 Gaps: 14
US-09-830-837-6 (1-1052) x US-09-000-016-3 (1-2539)
Qy 125 ArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSer----- 141
Db 745 CGTGGCTTACAGCGCGCCGCAAGCGCGGAGTCCGCGAAGCGGGAAC 804
Qy 142 -----AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSer 159
Db 805 CGCGCGGACCTTGACGTCCT-----GAACGCGGACCGCGTGG 843
Qy 160 ArgProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArg 179
Db 844 GACCCCGCACCA-----GACGCGTCCGACGCTGTGGAGCG 879
Qy 180 HisSerSerArgArgLeuLeuArg----- 187
Db 880 CGTCACCAACGCGACCGGACCGCTCCGCGCATGCCCGGCGGCGGTCG 939

```
QY 188 ---AlaIleProArgGlnValAlaGlnThr-LeuGlnAlaAspValLeuTrpGlnMetG1 206
Db 940 CAGGGCCGCGCTCGACACGCTCCGTCGGGCAGATCGGCGCCCAAGGCGTGGTCCGCGCG 999
QY 206 yTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGlyLysHisPr 226
Db 1000 CTACGACGGCAAGGGCTGAAGATCGCGCTCGACACCGGTGTGACACAGCCATCC 1059
QY 226 oHisPheLys---AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAs 245
Db 1060 GGACCTGAAGGGCGGGTGACCGCTCCAAGAACTTCACCGCGCGCGCGCGCGCGCA 1119
QY 245 pGlyLeuGlyHisGlyThrPheValAlaGlyValIleAla-----SerMe 260
Db 1120 CAAGTGGGGCAGCGCACCCACCGCTCGCTCGCGCGCGCGCGCGCGCGCGCGCA 1179
QY 260 tArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAs 280
Db 1180 GGCAAGTACAGGGGTGTCACCGCGCGCGCGCGCTCAACGGCAAGGTCTCTCGACGA 1239
QY 280 nAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLy 300
Db 1240 CTCCGGTTTCGGCGACGACTCCGGCATCTTCGCGCGCATGGAGTGGCGCGCGCGCAGG 1299
QY 300 sIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheValas 320
Db 1300 CGCGGAGTGTCCATCAGCATGAGCTGGCGGC-----ATGGACACACCGGAGACCGA 1350
QY 320 p-----LysValTrpGluLeuThrAla---AsnAsnValIleMetValSerAl 335
Db 1351 CCGCTGGAGCGCGGTGCACAAAGCTGTCGCGCGAGAGGGCGTCTGTTCGCCATCGC 1410
QY 335 aIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspVa 355
Db 1411 GGCGGCAACGAGGGCCG-----GAGTCGATCGTTCGCGCGCGCGCGCGCGCGCGC 1464
QY 355 lIleGlyValGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMe 375
Db 1465 CCTCACCCTGGCGCGGTGACGACACAGGACAAAGTCCGCGACTTCTCTCCACCGGCC 1524
QY 375 tThrThrTrpGluLeuProGlyTyrGlyArgMetLysProAspIleValThrTyrG1 395
Db 1525 CCGCCTC-----GGCGACGGCGCCATCAACCGGAGCTACCGCTCCCGC 1569
QY 395 yAlaGlyValArgGlySerGlyValLysGly----- 405
Db 1570 CFTGGACATCACGGCGCTCGCGGAGGGCAACGACATCGCGCAGGAGTGGTGGGG 1629
QY 406 -----GlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAl 423
Db 1630 ACCGGCGGCTACATGACCATCTCCGCGACGTGATGGCGACCGCGCGCGCGCGCGC 1689
QY 423 aValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLy 443
Db 1690 GCGGCGCTCTG-----AAGCAGCAGACCCCGACTGGACCTCGCGCGCAACTGAA 1740
QY 443 sGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGluGlnGlyHi 463
Db 1741 GGGCGGCTCACCGGCTCCACCAAGGCG---GGCAAGTACACCCCGTTCGAGCAGGGTTC 1797
QY 463 sGlyLysLeuAspLeuLeuArgAlaTyrGln 473
Db 1798 GGGCGGATCCAGGCGGACAAAGCGCTCCAG 1828
```

Search completed: June 2, 2003, 05:46:17
Job time : 337 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 01:15:11 ; Search time 535 Seconds
(without alignments)
4428.228 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNIVLLVLLCGKKH.....PRVKPQLMQVHPKPTPSV 1052

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q/cgr2_1/USFTO.spool/US09830837/runat_23052003_181923_8059/app_query.fasta_1.1223
-DB=N_Geneseq_101002 -QFIX=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837@cgrn_1.1.344 @runat_23052003_181923_8059 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5617	100.0	4338	21	Human subtilisin-k
2	5617	100.0	4338	24	Human site-1 prote
3	5487	97.7	4198	24	Hamster site-1 pro
4	5476	97.5	3895	21	Rat subtilisin-kex
5	5427	96.6	3788	21	Mouse subtilisin-k
6	3560	63.4	2729	24	Human polynucleoti
7	2366.5	42.1	3160	22	Drosophila site 1
8	2231	39.7	2859	23	Drosophila melanog
9	2219.5	39.5	5332	23	Gastric cancer ass
10	1229.5	21.9	786	20	Human gene express
11	1155	20.6	757	20	Human polynucleoti
12	474.5	8.4	469	22	Human excretory re
13	348	6.2	9689	22	Human kidney relat
14	348	6.2	9689	22	Coding region of t
15	341	6.1	1220	9	apr gene encoding
16	341	6.1	1220	11	Sequence of the ap
17	341	6.1	1524	7	DNA encoding the a
18	339	6.0	2044	9	Subtilisin gene fr
19	338.5	6.0	1500	8	B subtilis subtili
20	338.5	6.0	1500	16	Bacillus subtilis
21	338.5	6.0	1500	20	Subtilisin gene.
22	338	6.0	1499	11	Bacillus subtilis
23	338	6.0	2760	20	Bacillus subtilis
24	337.5	6.0	1180	22	Subtilisin E.DNA,
25	337.5	6.0	1500	8	Bacillus subtilis
26	336	6.0	266	20	Human gene express
27	335.5	6.0	1140	24	Bacillus lichenifo
28	334	5.9	1220	8	Thermotable Bacil
29	332.5	5.9	1146	24	Bacillus subtilis
30	331.5	5.9	1494	24	Bacillus amyloliqu
31	331	5.9	1497	18	Bacillus amyloliqu
32	331	5.9	1497	18	Bacillus amyloliqu
33	331	5.9	1497	20	Bacillus amyloliqu
34	331	5.9	1497	20	B. amyloliquefacie
35	331	5.9	1497	20	B. amyloliquefacie
36	331	5.9	1497	20	Subtilisin DNA seq
37	331	5.9	1497	21	Bacillus amyloliqu
38	331	5.9	1497	24	Bacillus amyloliqu
39	331	5.9	1497	24	Bacillus amyloliqu
40	330	5.9	1496	8	Streptomyces virid
41	330	5.9	2539	18	Dhpa-mel chimeric
42	330	5.9	2809	18	B. amyloliquefacie
43	329.5	5.9	1492	16	Subtilisin gene.
44	328.5	5.8	1494	20	B. amyloliquefacie
45	328.5	5.8	1494	20	B. amyloliquefacie

ALIGNMENTS

RESULT 1
AAA57197
ID AAA57197 standard; cDNA; 4338 BP.
XX
AC AAA57197;
XX
DT 03-OCT-2000 (first entry)
XX
DE Human subtilisin-kexin isoenzyme 1 cDNA.
XX

Human; subtilisin-kexin isoenzyme 1; SKI-1; antilipase; cytostatic;
vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
Ras-dependent cancer; restenosis; amyloid protein formation;
pro-brain-derived neurotrophic factor; proBDNF;
steroid-regulatory element-binding protein; SREBP; ds.
Homo sapiens.
XX
OS
XX


```
QY 461 GlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
DB 1877 CAAGGCCACGGCAAGCTCGCTCAGAGCCTATCAGATCCTCAACAGCTACAGCCA 1936
QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
DB 1937 CAGGCAAGTTTGAGCCCAAGCTACATAGATCTGACTAGTGTGCTTACATGTGGCCCTAC 1996
QY 501 CysSerGlnProIleTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
DB 1997 TGCCTCCACCCCATCTACTATGAGGAATGCCACAGTTGTAAATGTCACCATCCTCAAC 2056
QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
DB 2057 GGCATGGGAGTCACAGGAAGATTGTAGATAAGCCTGACTGCGAGCCCTATTITGCCACAG 2116
QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly 560
DB 2117 AACGGAGACAACATTGAAGTTGCCTTCTCTACTCTCGGTCTTATGGCCTTGGTCGGGC 2176
QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
DB 2177 TACCTGGCCATCTCCATTTCTGTGACCAAGAAGCGCTTCTTGGGAAGCATTTGCTCAG 2236
QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
DB 2237 GGCCATGTCATGATCCTGCTGCTCCCGCAGCAGACAGAGTCACAAAATGGTGCAGAA 2296
QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSer 620
DB 2297 CAGATTCACAGTAAGCTCCCATTAAGGTGAAGATAATCTACTCCCGCGCAAGC 2356
QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg 640
DB 2357 AAGAGAGTTCTCTGGGATCAGTACCACCAACCTCCGCTATCCACTGCTTATTTCCCGAG 2416
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
DB 2417 GATAATTTAAGGATGAAGAATGACCTTTAGACTGGAATGGTGATCACCACCAAT 2476
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
DB 2477 TTCAGGGATATACACAGATCTGAGAAGCATGGGCTACTTTGTAGAGGTCTCTGGGGCC 2536
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
DB 2537 CCCTTCACGTGTTTGTATCCAGTCAGTATGACACCTTCTGATGGTGACAGTGAGGAG 2596
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu 720
DB 2597 GAGTACTTCCCTGAAGAGATCGCAAGCTCCGGAGGAGCTGACAAACGGCCTCTCGCTC 2656
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
DB 2657 GTCATCTTCAGTGACTGGTACACACTCTCTTATGAGAAAAGTGAAGTTTATGATGAA 2716
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu 760
DB 2717 AACACAGGAGCTGGTGGATGCGGATACCGGAGGAGCTACATCCAGCTCTGAAATGAG 2776
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
DB 2777 CTGCTCTCTGTGGAAACATGGGTTTCAGCGATGGCTGTATGAAGGGAGTTCACCTG 2836
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
DB 2837 GCCAACCATGACATGTATTATGCTCAGGGTCAGCATCGCAAGTTTCCAGAAGATGCG 2896
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLysGlnGluThrAla 820
DB 2897 GTCGTGTAACACAGACTTTCAGGACCAAGGATTGGAGGTTTAAAGCAGGAACAGCA 2956
```

```
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
DB 2957 GTTGTGTTGAAACGTCCTCCCATTTTGGGACTTTATCAGATTCAGCTGAGGGTGGAGCCGG 3016
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
DB 3017 ATTGTACTGTATGGGACTCCAATTGCTTGGATGACAGTACCAGACAGAGGACTGCTTT 3076
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
DB 3077 TGGCTTCTGGATGCCCTCTCTCCAGTACATCGTATGGGTGACACCGCTAGCTCAGT 3136
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
DB 3137 CACTCTGGGAACCGCCAGCGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG 3196
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
DB 3197 GAAGGAACCATCTTTCATCGTACTCCAAGTTCTGGAGGCCATTTGGAGACCCCAAA 3256
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
DB 3257 CCTCGGCCCTTACCAGCCTGTCCACGCTTGTCTTGGGCCAAGCCAGCCCTTTAAGCAG 3316
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
DB 3317 ACGGCCCCAGTAGTAACCTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGTG 3376
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
DB 3377 GTGTTTACCACACTTTCGATCGAATCGCCTCAAGTAGGGCCCTTGTCCTCGGAGAGC 3436
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
DB 3437 GCGCCTGGACATCTCTGGAGGATCATGCTTGGCCCTACAAACAGGAGGTGGCCAG 3496
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGln 1020
DB 3497 ACCATTCTGTCTTTGCTTCCCTGGAGCCATGGTGTCTCTGGCTTCTTTGTGTACAA 3556
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
DB 3557 ATCAACAGGCCCAAGAGCAGGCCGAAGCGGAGGAGGCCAGGCTGAAGCGCCCGCAGCTC 3616
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
DB 3617 ATGCACAGGTTTCCCGCCCAAGACCCCTTCGGTG 3652
```

RESULT 2

AAD24182
ID AAD24182 standard; DNA; 4338 BP.

AC AAD24182;

XX 07-MAY-2002 (first entry)

XX Human site-1 protease DNA.

DE DE

KW Site-1 protease; SLP; promoter; transcription; obesity; type II diabetes;
KW sterol regulatory element-binding protein; SREBP; hypercholesterolaemia;
KW dyslipidaemia; atherosclerosis; cardiovascular disease; human; ds.

OS Homo sapiens.

XX Key Location/Qualifiers

PH CDS 497..3655

FT /*tag= a

FT /product= "Site-1 protease"

XX WO200200873-A1.

XX 03-JAN-2002.

XX

400/1000

CC The invention relates to human site-1 protease (SLP) promoter region.
CC The promoter sequence is useful for identification of compounds that
CC inhibit transcription of SLP, which in turn results in inhibition
CC of sterol regulatory element-binding protein (SREBP) pathway. The
CC compound identified is useful for the treatment of medical conditions
CC related to obesity, type II diabetes, hypercholesterolemia,
CC dyslipidemia, atherosclerosis and other cardiovascular diseases.
CC The present sequence is hamster SLP DNA.

XX
SQ Sequence 4198 BP; 1061 A; 1014 C; 1106 G; 1017 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 4198
Score: 5487.00 Matches: 1021
Percent Similarity: 98.57% Conservative: 16
Best Local Similarity: 97.05% Mismatches: 15
Query Match: 97.69% Indels: 0
DB: 24 Gaps: 0

US-09-830-837-6 (1-1052) x AMD24185 (1-4198)

QY 1 MetLysLeuValAsnLlePrLLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
DB 387 ATGAGCTCATCAACATCTGGCTTCTTCTGTGTGTTTCTCTGGAAAGAGCAT 446
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
DB 447 CTGGGTGACAGCTGGGGAGAAAGCGTTGAAAGGCAATCATGCCCTAGCTTCTCCAC 506
QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrLleValAlaPhe 60
DB 507 CTGACTTTGAAGTGGAAATCTCCCTCAACTGTGTGGAATATGAATATATATGTGGCTTTC 566
QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
DB 567 AACGGATATCTACAGCCAAAGCTAGAACTCATATTTTCAAGTGTCTGCAAAAGCAGT 626
QY 81 GluValAspAsnTrpArgLleIleProArgAsnProSerSerAspTyrProSerAsp 100
DB 627 GAAGTAGACAACCTGGAGATTTATACCTCGGAACAACCCATCCAGTACTACCTAGTAT 686
QY 101 PheGluValLleGlnLleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
DB 687 TTTGAGTGATTACAGATAAAGAGAAGCAGAGGCCGGCTGCTCACACTTGAAGATCAT 746
QY 121 ProAsnLleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
DB 747 CCACACATCAAGCGGTGACACCTCACCCAAAGTCTTTCTGTCTTGAGTGTTCGTA 806
QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
DB 807 TCTGACCCCATTTGGCCATGTATGAACCTCGGTGGAGCCAGAGTGGCAGTCAACGA 866
QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
DB 867 CCCTGGAAGAGAGCCAGTCTCTCCCTGGGCTCTGGATTCTGGCATGTCAACAGGAAGACAT 926
QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
DB 927 TCAAGCCGCGATTGTGTAGAGCCATTCTCGACAGAGTTGCCAGACATTTGCCAGCAT 986
QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
DB 987 GTGCTGTGGCAGATGGATACACAGTGCTAATGTCTCAGGTTGCTGTTTGTATCTGGG 1046
QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
DB 1047 CTCAGTGAGAGACATCCACACTTCAGAAATGTGAAGAGAGAACCACTGGACCAATGAG 1106
QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValLleAlaSerMet 260
DB 1107 CGACCCCTGGATGATGGCTGGCCATGGCACATTTGTCCAGGTGTGATTGCCAGCATG 1166

QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
DB 1167 AGGGAGTCCAGGAGATTGCCACAGATCCAGAGTGCACATCTTCGGGTCTTTACCAAC 1226
QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLys 300
DB 1227 AATCAGGTGTCTTACACATCTTGGTGTGGACGCTTCACTATCCATCTCCTAAAGAG 1286
QY 301 IleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheValAsp 320
DB 1287 ATTGATGTTCTAAACCTTAGCATCGCGGGCTGACTTCATGGATCATCCCTTTCTTAC 1346
QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
DB 1347 AGGTGTGGGAATTAACAGCTAACATGTATCATCTGTTCTCTATCGGCATGATGGA 1406
QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGly 360
DB 1407 CCTCTTATGGCACTCTGAATACCCAGCTGATCAGATGGATGTGATTGGATGGGTGGC 1466
QY 361 IleAspPheGluAspAsnLleAlaArgPheSerSerArgGlyMetThrThrTrpIleuLeu 380
DB 1467 ATTGACTTTGAAGATAACATCGCGCGCTTCTTCCAGGGGAATGACTACCTGGGAATA 1526
QY 381 ProGlyGlyTyrGlyArgMetLysProAspLleValThrTyrGlyAlaGlyValArgGly 400
DB 1527 CCAGAGGCTATGTGTCGCGAAACCTGACATGTGACCTATGCTGCTCCGAGTGGGGT 1586
QY 401 SerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
DB 1587 TCCGGTGTGAAGGGGCTCGCGGCACCTCTCAGGGACCACTGCTGCTTCCCGCATGGTT 1646
QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
DB 1647 GCTGGGCTGTCACTTGTGTAGTAAAGCAGTGCAGAGCGGAGCTAGTGAATCTCTGCC 1706
QY 441 SerMetLysGlnAlaLeuLleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460
DB 1707 AGTGTGAAGCAAGCCCTGATTGTCATCAGCCGGAGGCTCTCTGCTGTAACTATGTCAG 1766
QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnLleLeuAsnSerTyrLysPro 480
DB 1767 CAAGGCCATGCAAGCTGATCTGTCGAGCCTATCAGATCTCAGCAGCTACAAACCA 1826
QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
DB 1827 CAGGCGACTTGACTCTAGTACATCGACCTGACTGAGTGCCTTACATGTGGCTTAC 1886
QY 501 CysSerGlnProLleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
DB 1887 TGTCTCAGCCATCTACTATGAGGAATGCCAAATGTATGTACCTCAGTCTTATCTCAAT 1946
QY 521 GlyMetGlyValThrGlyArgLleValAspLysProAspTrpGlnProTyrLeuProGln 540
DB 1947 GGCATGGGAGTCAACAGGAAGATTTGTGGATAAGCCTGAGTGGCGGCCCTTATTACACAG 2006
QY 541 AsnGlyAspAsnLleGluValAlaPheSerTyrSerSerValLeuTrpProTyrSerGly 560
DB 2007 AATGAGACAACATTAAGTGGCTCTCTACTCTCTCAGTGTATGGCTTGTGGTCAAGC 2066
QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluLylleAlaGln 580
DB 2067 TACCTGGCATCTCCATTTCTGTGACCAAGAGCAGCTTCTCTGGAAGGATTCACAG 2126
QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
DB 2127 GCTCACATCATGATCAGGTGGCTTCTCCAGCAGAGAGCGAAGCAAAATGTTGTCGAG 2186
QY 601 GlnThrSerThrValLysLeuProLleLysValLysLleIleProThrProProArgSer 620
DB 2187 CATCTTCCAGTGAAGCTTCCATTAAGGTGAAGATCATCTCCACCCCTCTCTCGGAGC 2246
QY 621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg 640

Db 2247 AAGAGAGTCTCTGGGACCAAGTATCACAACTCCGCTACCCCGCAGGCTACTTTCCAGG 2306
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisIleHisThrAsn 660
Db 2307 GACAACCTTCGGGATGAGATGATCTTTAGACTGGAATGGCCAGCATGTCCACACCAAT 2366
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2367 TTCAGGACATGTACCAGCACCTTCGCAGCATGGCTACTTCTGTGAGGTGCTCGGTGCC 2426
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
Db 2427 CCATTTCACGTGCTTTGATGCTACACAGTATGGCAGCTTTCCTCATGTGGATGTAAGAA 2486
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
Db 2487 GAGTACTCCAGAGGAGATGGCAGCTGAGGAGGAGCATGCCAATGGCCCTTCCCTC 2546
QY 721 ValIlePheSerAspTyrPyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2547 GTCATCTTCAGTGACTGGTACAACACTTCTGTATGAGAAAGTGAAGTTTACGATGAA 2606
QY 741 AsnThrArgGlnTyrPyrMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2607 AACACAAGCCAGTGGTGGATGCCAGATCTGGAGGAGCCCAACATCCAGCTCTGAACGAG 2666
QY 761 LeuLeuSerValTyrAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
Db 2667 CTGCTGCTGTGTGGAAACATGGGTTCAGCGATGGCCCTTATGAGGGGAGTTGGCCCTG 2726
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2727 GCGAATCATGACATGATATTATGATCGGATCGGATCGAGCATGCCAAGTTTCCAGAAGATGT 2786
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2787 GTTGTGATCACACAGACTTTCAGGAGCAAGGATGGAGTCTTAAACAAGAGACAGCA 2846
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
Db 2847 GTTGTGAAATGTTCCATTTTGGGGCTTTATCAGATCCAGCTGAAGTGGGGCCGG 2906
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
Db 2907 ATCCTGCTGTATGAGATTCCAATTCCTGGATGACAGTCACAGACAGAAGGATGCTTT 2966
QY 861 TrpLeuLeuAspAlaLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2967 TGGCTTCTGGATGCACCTCTTCAGTACATCATATGGCGTGAACCTCCAGCCCTCAGC 3026
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3027 CATTCAGGACCGGAGCGCCAGCCAGCTGGAGCTGGCTGGCCCTCTCTGAAGAGATG 3086
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3087 GAAGGAACACCATCTCATGATCTCAAGGTCTTGGGCCCATCTGGAGACCCAAAA 3146
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3147 CCTCGGCCCTTCCAGCCCTGTCACACTTGTATGGGCCCAAGCCAGCCAGCTTTGAATGAG 3206
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3207 ACGGCCCCAGTAAATCTTGGAAACATCAGAGCTGCTCTCCATTGACCTGGCAAGTA 3266
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCAACTTTCATCGAATCGCCCTCAAGTGAGACCTTTGTCCTCGGAGAAAGT 3326
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000

Db 3327 GGTGCTGGGACATTCCTGGAGGATCATGCTCGTGGCGCTACAAACCAAGAGGTGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3387 ACCATCCCTGCTCTTGGCTTCTTCGGAGCCATGGTGGCCCTGGCCTTCTTTGGGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGGCCAAAGCCGCCGAGGAGGAGGCCAGGCGAAAGCGTCCACAGCTT 3506
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCACCACCACCAAGGACCCCGTCAGTG 3542
RESULT 4
AAA57198
ID AAA57198 standard; cDNA; 3895 BP.
XX
AC AAA57198;
XX 03-OCT-2000 (first entry)
DT
XX Rat subtilisin-kexin isoenzyme 1 cDNA.
DE
XX Rat; subtilisin-kexin isoenzyme 1; SKI-1; antilipemic; cytostatic;
KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
KW Ras-dependent cancer; restenosis; amyloid protein formation;
KW pro-brain-derived neurotrophic factor; proBDNF;
KW sterol-regulatory element-binding protein; SREBP; ds.
XX
OS Rattus sp.
XX
Key Location/Qualifiers
FH 418..3576
FT CDS
FT /*tag= a
FT /product= "SKI-1"
XX
PN WC200026348-A2.
XX
PD 11-MAY-2000.
XX
PF 04-NOV-1999; 99WO-CA01058.
XX
PR 04-NOV-1998; 98CA-2249648.
XX
PA (RECL-) INST RECH CLINIQUES MONTREAL.
XX
PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX WPI; 2000-365601/31.
DR P-PSDB; AAB06335.
XX
XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
PT producing a polypeptide useful for treating hypercholesterolemia, liver
PT steatosis and amyloidosis, comprises a specific amino acid sequence -
XX
PS Example 1; Page 61-66; 119pp; English.
XX
CC The present sequence encodes rat subtilisin-kexin isoenzyme 1 (SKI-1),
CC a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
CC from rat adrenal glands by RT-PCR using active site degenerate
CC primers. SKI-1 cleaves at a specific threonine residue within
CC the N-terminal segment of pro-brain-derived neurotrophic factor
CC (proBDNF). It is also capable of cleaving sterol-regulatory
CC element-binding proteins (SREBPs), which function to control lipid
CC biosynthesis and uptake in animal cells. Peptides which bind to and are
CC cleaved by SKI-1 may be used for monitoring SKI-1 activity, for screening
CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
CC activity. Proteic fragments of SKI-1 which bind to the SKI-1
CC catalytic site may be used as inhibitors of SKI-1 activity. They may
CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
CC substrate. Such diseases include hypercholesterolemia, high levels of
CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,

QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
|||||
Db 2398 TTCAGGACATGTACCAGCATCTGCGCAGCATGGGCTACTTTGTGGAGGTGCTTGGTGC 2457
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
|||||
Db 2458 CCATTACATGTTTGGACCCAGCGCAGTAGTACGCCTCTCTTATGTTGACAGTAGAGAA 2517
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgAtgAspValAspAsnGlyLeuSerLeu 720
|||||
Db 2518 GAGTACTTCCCTGAGGAGATTGCTAAGCTGAGGAGGAGCGTGACCAATGGCTTTCCCTT 2577
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
|||||
Db 2578 GTCGTCTTCAGTGACGTGTTACAACTCTCTGTTATGAGAAAGTGAAGTTTACGATGA 2637
QY 741 AsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu 760
|||||
Db 2638 AACACAAGGACAGTGGTGGATGCCAGATACCTGAGGAGGAGCCAGCTCCAGCTCTAAACGAG 2697
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrCluGlyGluPheThrLeu 780
|||||
Db 2698 CTGCTGCTGTGTGGAACTGGGTTCAGTGACGGCTGTATCAAGGGGAGTTGCCCTG 2757
QY 781 AlaAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
|||||
Db 2758 GCAAACCACGACATGTACTATGCATCGGGGTGCAGCATGGCCAGTTTCCAGAGATGT 2817
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
|||||
Db 2818 GTGGTGATCACACAGACTTTCAGGACCAAGGATTGGAAGTCTTAAACAAGACAGCA 2877
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
|||||
Db 2878 GTGTGCGAAGTATCCCAATCTCGGGCTATATCAGATTCCAGCTGAAGGTGGAGCCGG 2937
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
|||||
Db 2938 ATTGTGCTATGGAGACTCCAACCTGCTGGATGACAGTCACAGACAGAGACTGCTTT 2997
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
|||||
Db 2998 TGGCTTCTGGATGCACTCTTCAGTACACATCTATGTGTGACCCCTCCAGCCCTCAGC 3057
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
|||||
Db 3058 CATTCAGGAACCGGAGCGCCACCCAGCGGGGTGGCTGGCCCTCTCTGAAGGATG 3117
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
|||||
Db 3118 GAAGGAAACCACTTCATCGCTACTCCAAAGTCTTGGGCCACTTGGGAGACCGGAA 3177
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
|||||
Db 3178 CCTCGGCCCTCTCCAGCCTGTCCACACTTGTGTGGGCCAAGCCACAGCCTTTGAATGAG 3237
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
|||||
Db 3238 ACGGCACCCAGTAATCTTGGAAACACCAAGAGCTGCTCTCCATTGACCTGGCAAAAGTA 3297
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
|||||
Db 3298 GTGTTACCAACTTTCGCTCAATCGCCCTCAAGTGAGACCTTTGTCCTCCCTGGAGAAAGT 3357
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
|||||
Db 3358 GGTGCTGGGACATTCCTGGAGGATCATGCTGCGCGCTACAAACAGGAAGTAGGCCAG 3417
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln 1020
|||||
Db 3418 ACAWCCCTGTTTTTCCCTTCCTTGGAGCCATGGTGGCCCTTCCTCGTGGTACAG 3477
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040

Db 3478 ATCAGTAAGCCCAAGAGCCGCGGAAGCGGAGGAGCCCAAGGCAAGCGTCCACAACCTT 3537
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
|||||
Db 3538 GCACAGCAGCCCAACCCCTGCAAGGACCCCGTCAGTG 3573
RESULT 5
AAA57199
ID AAA57199 standard; cDNA; 3788 BP.
XX
AC AAA57199;
XX
DT 03-OCT-2000 (first entry)
XX
DE Mouse subtilisin-kexin isoenzyme 1 cDNA.
XX
KW Mouse: subtilisin-kexin isoenzyme 1; SKI-1; antilipaeic; cytostatic;
KW vasotropic; SKI-1 inhibitor; hypercholesterolaemia; liver steatosis;
KW Ras-dependent cancer; restenosis; amyloid protein formation;
KW pro-brain-derived neurotrophic factor; proBDNF;
KW sterol-regulatory element-binding protein; SREBP; ds.
XX
OS Mus sp.
XX
XX Key Location/Qualifiers
FH CDS 8..3166
FT /*tag= a
FT /product= "SKI-1"
XX
PN WO200026348-A2.
XX
PD 11-MAY-2000.
XX
XX 04-NOV-1999; 99WO-CA01058.
PF
XX 04-NOV-1998; 98CA-2249648.
PR
XX (RECL-) INST RECH CLINIQUES MONTREAL.
XX
PI Seidah N, Chretien M, Marcinkiewicz M, Laaksonen R, Davignon J;
XX WPI; 2000-365601/31.
XX P-PSDB; AAB06336.
DR
XX Novel soluble proteic fragment of subtilisin-kexin isoenzyme for
PT producing a polypeptide useful for treating hypercholesterolemia, liver
PT steatosis and amyloidosis, comprises a specific amino acid sequence -
XX Example 1; Page 67-71; 119pp; English.
PS
XX The present sequence encodes mouse subtilisin-kexin isoenzyme 1 (SKI-1),
XX a type-1 membrane-bound proteinase. The nucleotide sequence was isolated
CC from mouse corticotropic cells by RT-PCR using active site degenerate
CC primers. SKI-1 cleaves at a specific threonine residue within
CC the N-terminal segment of pro-brain-derived neurotrophic factor
CC (proBDNF). It is also capable of cleaving sterol-regulatory
CC element-binding proteins (SREBPs), which function to control lipid
CC biosynthesis and uptake in animal cells. Peptides which bind to and are
CC cleaved by SKI-1 may be used for monitoring SKI-1 activity, for screening
CC inhibitors of SKI-1 activity, or for screening enhancers of SKI-1
CC activity. Proteic fragments of SKI-1 which bind to the SKI-1
CC catalytic site may be used as inhibitors of SKI-1 activity. They may
CC be used to treat diseases involving overexpression of SKI-1 or SKI-1
CC substrate. Such diseases include hypercholesterolaemia, high levels of
CC fatty acids, lipids or farnesyl pyrophosphate, liver steatosis,
CC Ras-dependent cancer, restenosis and amyloid protein formation.
XX
SQ Sequence 3788 BP; 915 A; 967 C; 1015 G; 891 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 3788

Score: 5427.00 Matches: 1008

Percent Similarity: 98.19% Conservative: 25
Best Local Similarity: 95.82% Mismatches: 19
Query Match: 96.62% Indels: 0
Dbs: 21 Gaps: 0

US-09-830-837-6 (1-1052) x AAA57199 (1-3788)

QY 1 MetLysLeuValAsnIleTrpLeuLeuValValLeuLeuCysGlyLysLysHis 20
|||||.....
DB 8 ATGAAGCTCGTACAGACCTGGCTTCTGTGCTGGTGTCTGTGGAAAGCGCAG 67
|||||.....
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
|||||.....
DB 68 CTGGCGACAGCTGGGAGAGAGCTTGGAAAGCGCGCTGCCACAGCTGCCAC 127
|||||.....
QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrlGluValAlaPhe 60
|||||.....
DB 128 CTGACTTTGAAGTGGAAATCTCTCAACTGTGTGGTAGCAATATATGTGGCTTC 187
|||||.....
QY 61 AsnGlyTyrlPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSer 80
|||||.....
DB 188 AACGGATCTTCACAGCCAAAGCTAGAACTCATTTATTTCAAGTGGCTGAAAGCAGT 247
|||||.....
QY 81 GluValAspAsnTrpArgIleIleProArgAsnProSerSerAspTyrlProSerAsp 100
|||||.....
DB 248 GAAGTGGAAATCTGGAGATAATACCTCGGAACAACCCATCCAGTACTACCTAGTAT 307
|||||.....
QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
|||||.....
DB 308 TTTGAGGTGATTCAGATAAAGAGAGACAGAGCGGGCTCTCACATTTGAAGATCAC 367
|||||.....
QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrlAlaGlu 140
|||||.....
DB 368 CCACATCAAGCGGTGACACCCCGCGGAGAGTCTTCGTTCCTCAAGTTTGCTGAA 427
|||||.....
QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
|||||.....
DB 428 TCCAAACCCATCGTCCCTGTGTAAGAAACCCGGTGGAGCAGAGTGGCAGTCAACAGT 487
|||||.....
QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
|||||.....
DB 488 CCCCTGAAAGAGCCAGTCTCTCCCTGGGCTCTGGATTCGGCATGCAACAGAGACAT 547
|||||.....
QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
|||||.....
DB 548 TCAAGTCGGCGATTGTGTGAGAGCCATTCTCGCCAGGTGCGCCAGACACTGCGAGCAGAT 607
|||||.....
QY 201 ValLeuTrpGlnMetGlyTyrlThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
|||||.....
DB 608 GTGCTGTGGCAGATGGGATACACAGGTGCTAATGTCAGAGTTGCTGTTTTGATAGTGG 667
|||||.....
QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
|||||.....
DB 668 CTCAGTGAAGAGCATCCGCATTTAAGANVTGAAGAGAGAACCACTGGACCAATGAG 727
|||||.....
QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
|||||.....
DB 728 CGACCCCTGGATGATGGGTAGCCATGGCAGCATCTGTTGACAGGTGTGATGGCAGATG 787
|||||.....
QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
|||||.....
DB 788 AGGGAGTCCCAAGGATTTGCTCCAGATGCGAGCTGCATCTTCAGGGCTCTTACCAAC 847
|||||.....
QY 281 AsnGlnValSerTyrlThrSerTrpPheLeuAspAlaPheAsnTyrlAlaIleLeuLysLys 300
|||||.....
DB 848 AATCAGGTGTCTACACATCTTGGTTTGTGGATGCCCTTCACTATGTCATCTCTTAAAGAG 907
|||||.....
QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320
|||||.....
DB 908 ATGGACGTTCTCAACCTTAGCATCGGTGGGCCGACCTTCTATGATCATCTCGTTTGTGAC 967
|||||.....
QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
|||||.....

DB 968 AAGGTGTGGGAATTAACAGCTAACAACTAATATTGTTCTGCTATTGGCAATGATGGA 1027
QY 341 ProLeuTyrlGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyValGlyCly 360
|||||.....
DB 1028 CCTCTATGGCACTCTGATACCTCTGCTGATCAGATGATGATGATGGATGGGTGGC 1087
|||||.....
QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeu 380
|||||.....
DB 1088 ATTGACTTTGAAGATACATCGCTCGCTTTCTCCAGGGAATGATACCTACCTGGGAATTA 1147
|||||.....
QY 381 ProGlyGlyTyrlGlyArgMetLysProAspIleValThrTyrlGlyAlaGlyValArgGly 400
|||||.....
DB 1148 CCAGAGAGCTATGCTGTGTAAGCTGACATTTGCTACCTATGCTGCTGAGTGGGGGT 1207
|||||.....
QY 401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
|||||.....
DB 1208 TCCGGTGTGAAGGGGGCTGCGTGCACCTCTCAGGGACCACTGCTGCTCCCGAGTGGTC 1267
|||||.....
QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
|||||.....
DB 1268 GCTGGGGCGCTCAGCTTTGTAGTAAAGCAGTACAGAAAGCGGAGCTGGTGAATCTGCC 1327
|||||.....
QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGlu 460
|||||.....
DB 1328 AGTGTGAAGCAAGCTTTGATAGCTCAGCGCGAGACTTCTGGGGTCAACATGTTCGAG 1387
|||||.....
QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrlGlnIleLeuAsnSerTyrlLysPro 480
|||||.....
DB 1388 CAAGTCTATGGCAAGTTGGATCTGCTGCGAGCTTATCAGATCTCAGCAGCTATAAACCG 1447
|||||.....
QY 481 GlnAlaSerLeuSerProSerTyrlLysLeuThrGluCysProTyrlMetTrpProTyrl 500
|||||.....
DB 1448 CAGGCAAGCTTGAGTCTAGTACATCGACCTGACTGAGTGCCTTACATGTGGCCCTAC 1507
|||||.....
QY 501 CysSerGlnProIleTyrlGlyMetProThrValValAsnValThrIleLeuAsn 520
|||||.....
DB 1508 TGCTCCCAAGCTATCTACTATGGAAGATGCCAATCGTTATGTATGCTACCTCAAT 1567
|||||.....
QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrlLeuProGln 540
|||||.....
DB 1568 GGCATGGGCGTACAGAGAAATGTGGTAAGCTGAGTGGGAGCCCTATTATTACCAAG 1627
|||||.....
QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrlSerSerValLeuTrpProTrpSerGly 560
|||||.....
DB 1628 AATGAGAGCAACATTAAGTGGCTCTCTCTACTCTCTCTGTTGTGGCCCTGTGTCAGGT 1687
|||||.....
QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
|||||.....
DB 1688 TACCTTGGCATCTCCATTTCTGTGACCAAGAGCAGCTTCTGCGAAGGCATCGCTCAG 1747
|||||.....
QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
|||||.....
DB 1748 GCGCACATCATGATCAGCTGGCTGCCCGCAGCAGACAGTGTACACAGTGGTGGGAG 1807
|||||.....
QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer 620
|||||.....
DB 1808 CACATTCACCGTGAAGCTGCCATCAGGTGAAGATCATTCACCCCTCTCTCGGAGC 1867
|||||.....
QY 621 LysArgValLeuTrpAspGlnTyrlHisAsnLeuArgTyrlProProGlyTyrlPheProArg 640
|||||.....
DB 1868 AAGAGAGTCTCTGGGACCAAGTACCAACCTCGCTACCCACTGGCTACTTCTCCAGG 1927
|||||.....
QY 641 AspAsnLeuArgMetLysAsnProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
|||||.....
DB 1928 GACAACTTGGCGATGAAGATGACCTTTTAGACTGGAAATGGCGACCACTGCCACCAAC 1987
|||||.....
QY 661 PheArgAspMetTyrlGlnHisLeuArgSerMetGlyTyrlPheValGluValLeuGlyAla 680
|||||.....
DB 1988 TTCAGGACATGTACAGCATCTGGCGAGCATGGCTACTTCTGCTGGAGGTGCTCGGGCGCC 2047
|||||.....
QY 681 ProPheThrCysPheAspAlaSerGlnTyrlGlyThrLeuLeuMetValAspSerGluGlu 700
|||||.....
DB 2048 CCATTCACATGTTTGAAGCCACACAGTATGGCACTTGTCTGTGGTGGACAGTGAAGAA 2107
|||||.....


```
QY 701 GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu 720
DB 2108 GAGTACTCCCTGAGGAGATGCTAGCTGAGGAGGATGTGGCAATGGCCCTTCCCTC 2167
QY 721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
DB 2168 GTCATCTTCAGTGACTGCTACACACTCTGTTATGAGAAAGTGAAGTTTATGATGAA 2227
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
DB 2228 AACACAGCAGCGTGGTGGATGCGACACCGGAGGAGCGCAACATCCAGCTCTGAATGAG 2287
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
DB 2288 CTGCTGCTGTGTGGAACTAGGGGCTCAGTGACGGCTATATGAAGGGAGTTGTCCGT 2347
QY 781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
DB 2348 GCAACCATGACATGTAATGCTGCTGGGTGCGAGCATGCCAAGTTTCCAGAAGATGGC 2407
QY 801 ValValIleThrClnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
DB 2408 GTGCTGATCACACAGACTTTCAGGACCAAGGATGGAGGTCTTAAACAAGACAGACGA 2467
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
DB 2468 GTTGTGGAATGTTCCCATTTTGGGGCTTATCAGATTCATCCATCGAAGTGGAGCGCG 2527
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
DB 2528 ATCGTCTGTATGAGACTCCACTGCTTGGATGACAGTCACAGACAGAGGACTGCTT 2587
QY 861 TrpLeuLeuAspAlaLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
DB 2588 TGGCTTCTGATGCGTCTCTAGTACATCCTATGCGGTGACCCCTCCACGCCCTCAGC 2647
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
DB 2648 CATTCAGGGAACCGGAGCGCCACCTAGCGGAGCGGCTTGGCCCTCTCTGAAGGATG 2707
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
DB 2708 GAAGGAACACCATCTCCATCGGTACTCCAAAGTTCTTGAAGCCCACTTGGAGACCGGAA 2767
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
DB 2768 CCTCGGCCCTGCGGCGCTGCCACATTTGTCATGGGCCAAGCCACAGCCTTTGAATGAG 2827
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
DB 2828 ACGGCACCCAGTAATCTTTGGAAACATCAGAGCTGCTCTCCATTACCTGGACAAGTA 2887
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
DB 2888 GTGTTTACCACAACTTTCATCCATCGCCCTCAAGTGAGAGACTTGTCCCTCGAGAGAGT 2947
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
DB 2948 GTGCTCTGGGACATTCCTGGAGGATCATGCTTGGCCGCTTACAAACAGGAGGTGGACAG 3007
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
DB 3008 ACCATCCCGCTCTTCGCTTCTCGGAGCATGCTGTCCTGCGCTCTTTGTGGTACAG 3067
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
DB 3068 ATCAGAAGGCCAAGACCGCGGCGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3127
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
DB 3128 GCACAGCAGGCCACCCCTGCAGAGGACCCCATCATG 3163
```

RESULT 6

ABL90664/c
ID ABL90664 standard; cDNA; 2729 BP.

XX AC ABL90664;

XX DT 24-MAY-2002 (first entry)

XX DE Human polynucleotide SEQ ID NO 1226.

XX KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
XX KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
XX KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
XX KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; human; secreted protein; gene; ss.

OS Homo sapiens.

XX PN WO200190304-A2.

XX PD 29-NOV-2001.

XX PF 18-MAY-2001; 2001WO-US16450.

XX PR 19-MAY-2000; 2000US-205515P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Birse CE, Rosen CA;

XX DR WPI; 2002-122018/16.

XX DR P-PSDB; ABB90255.

XX PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -

XX PS Claim 4; SEQ ID NO 1236; 2081pp + Sequence Listing; English.

XX CC The invention relates to novel genes (ABL9449-ABL90853) and proteins
CC (ABB9040-ABB9044) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.

XX CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at http://wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2729 BP; 642 A; 722 C; 674 G; 686 T; 5 other;

Alignment Scores:

Pred. No.: 6,72e-241 Length: 2729
Score: 3560.00 Matches: 665
Percent Similarity: 99.55% Conservatives: 0
Best Local Similarity: 99.55% Mismatches: 3
Query Match: 63.38% Indels: 0
DB: 24 Gaps: 0

US-09-830-837-6 (1-1052) x ABL90664 (1-2729)

QY 385 GlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLys 404
|||||

```
Db 2726 GGTCCGATGAACCTGACATTTCTACCTATGGTGTGGCGTGGCGGGTCTGCGCGTGA 2667
Qy 405 GlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValAlaGlyAlaVal 424
Db 2666 GGGGGGTGGGGGCTCTCAGGGACCAAGTGTTCCTCCAGGTGGTGGAGTGTCTGTC 2607
Qy 425 ThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLysGln 444
Db 2606 ACCTTGTGTAGTGACACAGTCCAGAAAGCGTGGTGGTGAATCCCGCAGTATGAAGCAK 2547
Qy 445 AlaLeuLeuAlaSerAlaArgArgLeuProGlyValAlaSerMetPheGluGlnGlyHisGly 464
Db 2546 GCCCTGATCGCGTCAGCCGGAGGCTCCCGGGGTCAACATGTTTGAGCAAGSCCAGGCG 2487
Qy 465 LysLeuAspLeuLeuArgAlaTyrGlnLeuAsnSerTyrLysProGlnAlaSerLeu 484
Db 2486 AAGCTCGATCTGTAGACCTATCAGATCTCAACAGCTACAGCCACACAGGCCAGTTTG 2427
Qy 485 SerProSerTyrIleAspLeuThrGluCysProTyrMetTyrProTyrCysSerGlnPro 504
Db 2426 AGCCCACTACATAGATCTGACGTAGTGCCTACTAGTGTGGCCTACTCTCCAGGCC 2367
Qy 505 IleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyVal 524
Db 2366 ATCTACTATGAGGAATGCCAGACTTGTAAATGTCAACATCTCTCAACGGCATGGAGTCT 2307
Qy 525 ThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGlyAspAsn 544
Db 2306 ACAGGAAGAATTGTAGATAAGCTGACTGGCAGCCCTATTTCACCAAGAACAGGAGCAAC 2247
Qy 545 IleGluValAlaPheSerTyrSerValIleTyrProTyrProTyrSerGlyTyrLeuAlaIle 564
Db 2246 ATTGAAGTGTCTCTCTACTCTCTGCTCTGCTTATGGCCTGTGTGGCTACCTTGGCCATC 2187
Qy 565 SerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisValMet 584
Db 2186 TCCATTTCTGTGACCAAGAAAGCGCTCTCTGGGAAGGCAATGTCTCAGGSCCATGTGATG 2127
Qy 585 IleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSerThr 604
Db 2126 ATCACTGTGGCTTCCCCAGCAGACAGAGTCAAAAATGTTGTCAGAACAGACTTCAACA 2067
Qy 605 ValLysLeuProIleLysValLysIleThrProThrProProArgSerLysArgValLeu 624
Db 2066 GTAAAGCTCCCATTAAGGTGAAGATAATTCCTACTCCCGCGGACAGACAGATTCTC 2007
Qy 625 TrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLeuArg 644
Db 2006 TGGGATCAGTACCAACCTCCGCTATCCACCTGGCTATTTCGCCAGKATATTTAAGG 1947
Qy 645 MetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAspMet 664
Db 1946 ATGAAGAATGACCTTTAGACTGGAAATGGTATCATCATCCACACCAATTTTCAGGGATATG 1887
Qy 665 TyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThrCys 684
Db 1886 TACCAAGCATCTGAGAAGCATGGGCTACTTTGTAGAGTCTCTCGGGGCGCCCTTCACGTGT 1827
Qy 685 PheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPhePro 704
Db 1826 TTTGTATGCCATCAGTATGGCACTTTGCTGATGTGGACAGTGGAGGAGGAGTACTTCCT 1767
Qy 705 GluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIlePheSer 724
Db 1766 GAAGAGATGCCAAGCTCCGGAGGAGCTGGACACAGCCCTCTCGCTCGTCATCTTCAGT 1707
Qy 725 AspTyrTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArgGln 744
Db 1706 GACTGTACACACTCTCTGTATGAAAAGTCAAGTCTTTATGATGAAAACACAGGAGCAG 1647
Qy 745 TrpTrpMetProAspThrGlyCysAlaAsnIleProAlaLeuAsnGluLeuLeuSerVal 764
Db 1646 TGGTGGATGGCGGATACCGGAGGAGCTAACATCCCAAGCTCTGAATAGAGTGTCTGTGTG 1587
Qy 765 TrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAsp 784
Db 1586 TGGAAACATGGGGTTTCAGCGATGGCTGTATGAAGGGAGTTCACCTCGCCCAACCATGAC 1527
Qy 785 MetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIleThr 804
Db 1526 ATGTATTATGCTGTCAGGGTGCAGCATCGCAAGTTTCAGAAAGTGGCTCTGTATAACA 1467
Qy 805 GlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGlnThrAlaValIleGluAsn 824
Db 1466 CAGACTTTCAGAGGACCAAGGATTCGAGCTTTTAAAGCAGGAAACACAGCTGTGTGAAAC 1407
Qy 825 ValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyGlyValIleValLeuTyr 844
Db 1406 GTCCCACTTTTGGGACTTTATCAGATTCACAGTCCAGCTGAGGTGGAGCGCGGATGTACTGTAT 1347
Qy 845 GlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheThrLeuLeuAsp 864
Db 1346 GGGGACTCCAAATGCTTTGATGACAGTCCCGACAGAGAGCTGCTTTTGGCTTCTGGAT 1287
Qy 865 AlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSerGlyAsn 884
Db 1286 GCCCTCTCCAGTACATCGTATGGGTGACACCGGCTAGCCTCAGCTCAGTCTGGGAAC 1227
Qy 885 ArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHis 904
Db 1226 CGCCAGCGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGGATGGAAGGAACCAT 1167
Qy 905 LeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeu 924
Db 1166 CTTTCATCGTACTCCAAGTTCCTGGAGGCCCATTTGGGAGACCCAAAACCTCGGCCCTTA 1107
Qy 925 ProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGluThrAlaProSer 944
Db 1106 CCAGCCTGTCCACGCTGTCTTGGGCCAAGCCACAGCCTTTAAACGAGAGCGGCCCACT 1047
Qy 945 AsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysValIleLeuProAsn 964
Db 1046 AACCTTTGGAAACATCAGAAAGTACTCTCCATTGACCTGGACAGAGTGTGTGTACCCAAC 987
Qy 965 PheArgSerAsnArgProGlnValArgProLeuSerProGlyLeuSerGlyAlaThrAsp 984
Db 986 TTTTCATCGATCGCCTCAAGTGAAGCCCTTGTCCCTCGAGAGAGCGGCCCTGGGAC 927
Qy 985 IleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIleProVal 1004
Db 926 ATTCCTGGAGGATCATGCTGCGCTACCAACGAGGAGTGGGCGGACCATTCCTGTC 867
Qy 1005 PheAlaPheLeuGlyAlaMetValValLeuAlaPheValIleGlnIleAsnLysAla 1024
Db 866 TTTCCTTCTCGGAGCCATGTTGTCTCTGCTGCTTCTTGTGTGTACAAATCAACAAGGCC 807
Qy 1025 LysSerArgProLysArgArgLysProArgValLysArgProGlnLeuMetGlnGlnVal 1044
Db 806 AAGAGCAGCGCGAAGCGGAGGAGCCAGGGTGAAGCGCCGCGCAGCTCATCGACGAGTT 747
Qy 1045 HisProProLysThrProSerVal 1052
Db 746 CACCCGCAAAAGACCCCTTCGGTG 723
RESULT 7
AAA89290
ID AAA89290 standard; cdna; 3160 BP.
XX
AC AAA89290;
XX
DT 28-MAR-2001 (first entry)
XX
DE Drosophila site 1 protease SLP homologue (dsIP) cdna.
XX
KW Fly; site 1 protease; SLP; dsIP; SREBP;
KW sterol regulatory element binding protein; transgenic animal;
```

KW animal model; lipid metabolism; transcription factor; cholesterol;
 KW obesity; insulin resistance; therapy; diagnosis; pesticide; ss.
 XX

OS Drosophila melanogaster.

XX Key Location/Qualifiers
 XX CDS 62..3040
 FT /*tag= a

XX WO200076308-A1.

XX 21-DEC-2000.

XX 08-JUN-2000; 2000WO-US15880.

XX 14-JUN-1999; 99US-0332522.

XX 15-MAR-2000; 2000US-0189700.

XX (EXEL-) EXELIXIS INC.

XX Costa MA, Doberstein SK, Elson S, Ferguson KC, Homburger SA;
 PI Ebens AJ, Keegan KP, Stout TJ;
 XX WPI; 2001-091292/10.

DR P-PSDB; AAB20015.

XX Novel invertebrate organism genetically modified to express or
 PT mls-express steroid regulatory element binding protein pathway protein
 PT used as model system for studying lipid metabolism and determining
 PT lipid content -

XX Claim 9; Page 79-80; 90pp: English.

XX The present sequence is that of Drosophila melanogaster DNA
 CC encoding dSIP (see AAB20015), a homologue of mammalian site 1
 CC proteases (SIP) that are involved in sterol regulatory element
 CC binding protein (SREBP) processing, cleaving SREBP at the luminal
 CC loop. SREBPs are transcription factors that activate genes
 CC involved in cholesterol and fatty acid synthesis, and are major
 CC mediators of insulin action in the liver. Alterations in SREBP
 CC function and expression are implicated in obesity and insulin
 CC resistance. The dS2P DNA was identified from an expressed sequence
 CC tag database search on the basis of homology to hamster S2P. The
 CC invention provides fly and nematode (Caenorhabditis elegans)
 CC polynucleotides (see AAB9287-90) and encoded proteins (see AAB20012-15)
 CC involved in the SREBP pathway. C. elegans and D. melanogaster
 CC animals genetically modified to express or mis-express these
 CC proteins are claimed. These genetically modified animal models have
 CC identifiable phenotypes that make them useful in assays for studying
 CC lipid metabolism, other genes implicated in lipid metabolism and
 CC compounds capable of modulating lipid biosynthetic pathways. Model
 CC organisms or cultured cells can be used to identify new drug
 CC targets, therapeutic agents, diagnostics and prognostics of
 CC disorders associated with lipid metabolism, and also to identify
 CC pesticide targets directed to components of the SREBP pathway.

XX Sequence 3160 BP; 985 A; 671 C; 667 G; 835 T; 2 other;

Alignment Scores:

Pred. No.: 7.53e-157 Length: 3160
 Score: 2366.50 Matches: 466
 Percent Similarity: 61.70% Conservative: 151
 Best Local Similarity: 46.60% Mismatches: 232
 Query Match: 42.13% Indels: 151
 DB: 22 Gaps: 16

US-09-830-837-6 (1-1052) x AAB9290 (1-3160)

QY 47 PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla 66

DB 116 TTTAAACAGCCGGTGTTCCTCAATGAGTTTCATCTCCACTTCCATCAAAATACCTTTGCC 175

QY 67 LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArg 86

DB 176 CCGGTCGGAATCCTACATCGCAGCAAAATCTTGGTTCAAACGTAAACGAAGTGGAGA 235
 QY 87 IleIleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIle 106
 DB 236 ATTGTTCCCGCTCTAAATTTGGCTTGGCAATATCCAGTAGTATTGATATCTTACGAGTT 295
 QY 107 LysGluLysGlnLysAlaGly-----LeuLeuThrLeuGluAspHisProAsn 122
 DB 296 TGCAGCGTTATGAATCATCATCAGAGTTTATTATAGAAAGGCTTCAGACTCCCATCA 355
 QY 123 IleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr---AlaGluSer 141
 DB 356 GTAAAGGCGAGTAGTTCCCGCAGCGAAGGATCCTAAACTATGACGCTATAGC 415
 QY 142 AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgPro 161
 DB 416 AACCTAACG----- 424
 QY 162 LeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSer 181
 DB 425 -----TATATTAC-----CGCCATCCC 442
 QY 182 SerArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThr 196
 DB 443 CAAGGAGTGTCTAAGGAACAGAAACCCCAACAGATCGCCACCGACAATTTGTCTCGTA 502
 QY 197 LeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaVal 216
 DB 503 CTCACGCCCAACATCTTTTGGAAAGCTGGGTATCACAGCAAGGAGCTAAAGTGGCCATT 562
 QY 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluAspGlyThrAsn 236
 DB 563 TTCACACTGGCCTAACCAAAACCATCCACACTTTCGAATGTAAAGGACGACGACAAAC 622
 QY 237 TrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyVal 256
 DB 623 TGCACGAATGAAAGTCACCTTGACGACAGAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 682
 QY 257 IleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArg 276
 DB 683 ATCGCTCTCTCCAGGAATGCCCTAGGCTTCGCTCCGACGCCCTTTTACATATTATAA 742
 QY 277 ValPheThrAsnAsnGlnValSerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAla 296
 DB 743 GTTTTACGAACTCCCAAGTTCTTACACTTCTGTTCTGTTCTGTTCTGTTCTGTTCTGTT 802
 QY 297 IleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHis 316
 DB 803 ATATATAGGAAATAAATCAATCTCAACCTTAGCATTTGGGGTCCCGACTTTATGGACTCG 862
 QY 317 PropheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIle 336
 DB 863 CCGTTCGTTGAAAGGTGTGGAACTGTCGGCTGAATAATGCTAATGATATGATATGCGGACGA 922
 QY 337 GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle 356
 DB 923 GGAATGATGTTCCCTTGTACGCGACGCTAAACAACTCTGGCGCATCAGACGATGATGTT 982
 QY 357 GlyValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThr 376
 DB 983 GCGTGTGGTGGCATTTCAGTTTGTATGATAAAATCCCAAGTTTAGTTCGAGAGGAATGACA 1042
 QY 377 ThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAla 396
 DB 1043 ACGTGGGAACCTTCCCTTAGCTACGCGACGCTATGGGACTCGATATTGTCAGTACGGAAGT 1102
 QY 397 GlyValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
 DB 1103 CAAGTGGAAAGCGAGTAGTGTGGCAAGGGGTGCAGACGACTCTCTCGAATACCTCGGTGCC 1162
 QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435

PA (PEKE) PE CORP NY.
 XX PI Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR P-PSDB; ABB68358.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions
 XX Claim 1; SEQ ID NO 31865; 21pp + Sequence Listing; English.
 PS The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2859 BP; 883 A; 608 C; 613 G; 755 T; 0 other;

Alignment Scores:

Pred. No.: 2-26e-147 Length: 2859
 Score: 2231.00 Matches: 445
 Percent Similarity: 59.20% Conservatives: 147
 Best Local Similarity: 44.50% Mismatches: 230
 Query Match: 39.72% Indels: 178
 DB: 23 Gaps: 17

US-09-830-837-6 (1-1052) x ABL12461 (1-2859)

QY 47 PheSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla 66
 DB TTTTAAACACGCGTGTTCCTCAATAGTTCATCGTTCACCTCCATCAAAATCTTGCC 114
 QY 67 LysAlaArgAsnSerPheIleSerAlaLeuLysSerSerGluValAspAsnTrpArg 86
 DB CGGTCGCGAAGTCCATACATCGCAGCAAACTCTTGGTTCAACAGTAAAGAACTGGAGA 174
 QY 87 IleIleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIle 106
 DB ATGTTCGCGTCAATTTGGCTTGGCAATATCCAGTGAATTTGATATCTTACGAGTT 234
 QY 107 LysGluLysGlnLysAlaGly-----LeuLeuThrLeuGluAspHisProAsn 122
 DB TGGCAGCGTTATGAATCATCATCAGAGTTTATTATAGAAAGGCTTCAGACTCACCCATCA 294
 QY 123 IleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr---AlaGluSer 141
 DB GTAAAGCGAGTAGTCCCGCAGCGAGCGTACGAAAGATCCTAACTATGACGCTATAGC 354
 QY 142 AspProThrValProCysAsnGluThrArgTyrSerGlnLysTrpGlnSerSerArgPro 161
 DB AACCTAAGC----- 363
 QY 162 LeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSer 181
 DB -----TATATTAC-----CGCCATCC 381
 QY 182 SerArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThr 196
 DB CAAAGGAGTGTAGAGAACACAAACCAACGATCGCCAGCAACAAATGTCTCCGTA 441
 QY 197 LeuGlnAlaSerValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaVal 216
 DB CTTCCAGCCCAACTCTTTGGAGAGTGGGTATACAGGCAAGGGAGTAAAGTGGCCATT 501
 QY 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn 236
 DB TTCGACACGTGGCTTACCAAAACCATCCACACTTTCGAATGTAAAGAACGACCAAC 561
 QY 237 TrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyVal 256
 DB TGGACGAATGAAAAGTCACTTGCAGCAGAGTCAGTCATGCGACCTTCGTCGCGCGG--- 618
 QY 257 IleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArg 276
 DB ----- 618
 QY 277 ValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAla 296
 DB -----GTTTCTACACTTCTCTGGTTCCTGGATGATGATCACTCACTACGCG 660
 QY 297 IleLeuLysLysIleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHis 316
 DB ATATATAGGAAATAAATCACTTCAACCTTAGCATTTGGGGTCCCGACTTATGGAGTCG 720
 QY 317 ProPheValAspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIle 336
 DB CCGTTCGTTGAAAAGGTGTTGGAACACTGTCGCTTAATAATCTCATATGATATCGGACGA 780
 QY 337 GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValle 356
 DB GGAATGATGGTCCCTTGTACGCGACGCTAAACAATCTCGCGATGAGCGATGTAGTT 840
 QY 357 GlyValGlyLysIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThr 376
 DB GCGTTCGTTGGTCATTCAGTTTGTATGATAAAATCCCAAGTTTATGTCGAGAGGAATGACA 900
 QY 377 ThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAla 396
 DB ACCTGGGAATTCCTTCCTAGCTACGACGATGGGACTCGATATGTCACGACTACGGAAGT 960
 QY 397 GlyValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAla 416
 DB CAAAGTGAAGGACGATGATGTGCAAGGGGTGCAGAGGACTCTCTGCAACATCCGTCTCC 1020
 QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435
 DB TCTCCAGTTTTCAGGGGCTGCTGCTCTTATAAGCGGTGCATTTTCAAGAAATCGAC 1080
 QY 436 LeuValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGly 455
 DB TACATAACCCACCATCTCTTAGCAGGTACTCATGAAGGTGCCGAGAACTGCCCAT 1140
 QY 456 ValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeu 475
 DB TATAACATGTTTGCAGCGGAGCTGGAAAACTGAATTTGCTGAAGAGTATGACGATTTG 1200
 QY 476 AsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysPro 495
 DB CTGTCATACAAACCAAGATAACCTTATTCGGCATACCTTGACTTCACCCAA---AAC 1257
 QY 496 TyrMetTrpProTyrCysSerGlnProIleTyrTyrGlyGlyMetProThrValValAsn 515
 DB TATATGGCCTTATAGCTCCCAACCTCTGTACTATGGAAGTCCGTCGCTATGCAAC 1317
 QY 516 ValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGln 535
 DB GTTACCATACTCAATGGTATCTCTGTCAACAAGTCATATAGTTGGCATCCCTAAATGGATT 1377
 QY 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeu 555
 DB CCGGATTTCGAAACCAAGGTCTTCTCAAGTATCTGCACAAGTTTCGCTATCGTT 1437
 QY 556 TrpProTrpSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrp 575
 DB TGGCCGTGGACCGGTGGATGTGAGTTTATTTGCTGTGTAAGAAAGGAGAGAAATTT 1497
 QY 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595

Db	1498	GAAGTGTTCGTAAGAGGAGTATACCTAGTTTGGAAAGCTTTAAACAGACACCAAC	1557	Db	2554	-----TGCAC	2559
Qy	596	LysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIleLeuPro	615	Qy	898	GluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGly	917
Db	1558	GAACCT-----CATGTTACAGAGTGCAGCTTTCCTTTTAAACAATAAAGGTTACTCCA	1608	Db	2560	-----AAATCAAGTTTATTGAGAAATCTAAATCGTATACTGAATTTACAAAATTA	2610
Qy	616	ThrProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProPro	635	Qy	918	AspProLysProArgProLeuProAla	926
Db	1609	AAACCGCAAGAACAGAGAGGATTTATGGGATCAGTACACAGCCTAAGGTATCCACCG	1668	Db	2611	GAGAGAGCACCATTACCCCTTAGGATATCGCAAGTATATATAAATCTCGTTCACAGGAC	2670
Qy	636	GlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp	655	Qy	927	-----CysProArgLeuSerTrpAlaLysProGlnProLeuAsnGluThrAlaProSer	944
Db	1669	CGCTATATCCACGAGATGATCTCAAGTTAACTAGATCTCTGGACTGGAGCGCAGAC	1728	Db	2671	ATAATTTGTGAACAATTTAAGTGGCTTCACCGACGACGAACAATAACGCC	2721
Qy	656	HisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal	675	Qy	945	AsnLeuTrpLysHisGlnLysLeuLeuSerLysLeuAspLysValLeuProAsn	964
Db	1729	CATATACACAACTTTAGGACATGTATACATTTACGAATGTGGCTACTACATT	1788	Db	2722	-----GAGGAAGGAATCTTCTATATATAGACGTACCATCTACTGGAAT	2766
Qy	676	GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet	695	RESULT 9			
Db	1789	GAITTTTTCGACAGACCTTCACCTGCTTCATGCTCGGATATGGCGGTATTGATT	1848	ABL12460/c			
Qy	696	ValAspSerGluGluTyrPheProGluIleAlaLysLeuArgArgAspVal---	714	ID	ABL12460	standard; cDNA; 5332 BP.	
Db	1849	GTTGACCTGAGAGAGGTTTGGCGAGGAGAAATAACGCTTTACAGGAAACAGTGAT	1908	XX	ABL12460;		
Qy	715	AspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerValMetArgLys	734	XX	26-MAR-2002	(first entry)	
Db	1909	AAAGAGGCTTGAATGCTGATTCGGAGACTGGTATACACCACTGTGTATGAAAAA	1968	XX	Drosophila melanogaster	expressed polynucleotide SEQ ID NO 31862.	
Qy	735	ValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsn	754	XX	Drosophila;	developmental biology; cell signalling; insecticide;	
Db	1969	ATTAAATCTTTCACGAGAACCCGACAAATGTGTGACACCCGACACTGGTGGCGCAAT	2028	XX	pharmaceutical; gene; ss.		
Qy	755	IleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyr	774	XX	Drosophila melanogaster.		
Db	2029	ATTCCAGCCTGAATGATTATTGAAGCATTTTGGAATGTCTTGGCATTTTTCGCT	2088	XX	WO200171042-A2.		
Qy	775	GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla	794	XX	27-SEP-2001.		
Db	2089	GAGGACATCTAAACTGGCGGACCATTCATCTACTATGCTAGTGGAGCCACATTTGT	2148	XX	23-MAR-2001; 2001WO-US09231.		
Qy	795	LysPheProGluAsp-----GlyValValIleThrGlnThrPheLysAspGlnGlyLeu	812	XX	23-MAR-2000; 2000US-191637P.		
Db	2149	AGTTTCCATGATCCAGGAGATATTATAGTGGGCACAAAACCTGAATGACCAAGACTT	2208	XX	11-JUL-2000; 2000US-0614150.		
Qy	813	GluValLeuLysGlnGluThr-----AlaValValGluAsnValProIleLeuGly	829	XX	(PEKE) PE CORP NY.		
Db	2209	TCGATTTATTCTTAAACACCCAGCAGGTAGCAAACTAGATGATGACCTATTTTGT	2268	XX	Venter JC, Adams M, Li PWD, Myers EW;		
Qy	830	LeuTyrGlnIleProAla-----	835	XX	WPI; 2001-658660/75.		
Db	2269	ATGTTCCAAACCAAGCGCAACAGTATTCAAAGCAACGAGAAATCGTGTCAATCGGAA	2328	XX	P-PSDB; ABB68357.		
Qy	835	-----	835	XX	New isolated nucleic acid detection reagent for detecting 1000 or more		
Db	2329	AGCAATTTGGCAGAGCTATACCCACAGATTACTCCACATTTAAGAACCGGTTTGTCTA	2388	XX	genes from Drosophila and for elucidating cell signalling and cell-cell		
Qy	836	-----	837	XX	interactions -		
Db	2389	CTGCAACGAGCAACGAGTATCAGTTTTCGAAAACCAATAATCATGAATAAGAAT	2448	XX	Claim 1: SEQ ID NO 31862; 21pp + Sequence Listing; English.		
Qy	838	GlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLys	857	XX	The invention relates to an isolated nucleic acid detection reagent		
Db	2449	GAAGGACGATTCGCTATATGGGACTCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2508	XX	capable of detecting 1000 or more genes from Drosophila. The invention is		
Qy	858	AspCysPheTrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProPro	877	XX	useful in developmental biology and in elucidating cell signalling and		
Db	2509	GCTTGTACTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2553	XX	cell-cell interactions in higher eukaryotes for the development of		
Qy	878	SerLeuSerHisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrPro	897	XX	insecticides, therapeutics and pharmaceutical drugs. The invention		
				XX	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
				XX	sequences (ABL01840-ABL16175) and the encoded proteins		
				XX	(AB57737-AB72072).		
				XX	The sequence data for this patent did not form part of the printed		
				XX	specification, but was obtained in electronic format directly from WIPO		
				XX	at ftp.wipo.int/pub/published_pct_sequences.		
				XX	Sequence 5332 BP; 1653 A; 998 C; 1004 G; 1677 T; 0 other;		
				XX			
				XX	Alignment Scores:		
				XX	Pred. No.: 3.22e-146	Length: 5332	
				XX	Score: 2219.50	Matches: 482	
				XX	Percent Similarity: 53.66%	Conservative: 163	

Best Local Similarity:	40.10%	Mismatches:	271
Query Match:	39.51%	Indels:	287
DB:	23	Gaps:	21
US-09-830-837-6 (1-1052) x ABL12460 (1-5332)			
QY	47	PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla	66
DB	4278	TTTAAACACAGCCGTTGTCACATGAGTTTCATCGTTCACTTCCATCAAAATACTTTGCC	4219
QY	67	LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluVal	82
DB	4218	CCGGTCCGAGATCCCTACATCGCAGCAAACTTCTTGGTCAACAGTAAAGTTGTATTCA	4159
QY	82	-----	82
DB	4158	AATGGTTATATTATACCTTAAATGCTATTCTCCTGTTTATTTCAACCGGATTGAGCCT	4099
QY	82	-----	82
DB	4098	GCTTCTTTTGAAGACACAAACATCTTTTCTGCAGAGAAAGTACCTTTTAAAGTTT	4039
QY	83	-----	83
DB	4038	AAACACTTTTCAACAGCTCTTTATATCCCTTTCCATTTTATAGTAACGAATCGGAGAT	3979
QY	87	eIleProArgAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLy	107
DB	3978	TGTTCCCGCTAAATTTGGCTGGCAATATCCAGTATTTTGATATCTTACGAGTTTG	3919
QY	107	sGluLysGlnLysAlaGly	123
DB	3918	CGACGGTTATGATCATCATCAGAGTTTATATAGAAAGCTTCAGACTCACCCATCAGT	3859
QY	123	eLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr	142
DB	3858	AAAGCAGTAGTTCCTCCACGAGCGTACGAGGATCTTAAACTATATGACGCTATAGCA	3799
QY	142	pProThrValProCysAsnGluThrArgTrpSerGlnLysTyrpGlnSerSerArgProLe	162
DB	3798	CCTAAGC	3792
QY	162	uArgArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSe	182
DB	3791	-----TATATTAC-----CGCATCCCCA	3772
QY	182	rArgArgLeuLeuArgAlaIlePro	197
DB	3771	AGGAGTGCTAAGAACAGAAACCCAAACACGATCGCCACGACAAATTGCTCCGTACT	3712
QY	197	uGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPh	217
DB	3711	CCACGCCACATCCTTTGGAAGCTGGGTATACAGGCAAGGAGTTAAAGTGGCCATTTT	3652
QY	217	eAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTr	237
DB	3651	CGACACTGGCTTAACCAAAACCATCCACACTTTCGAATGTAAAGGAACGAACAACTG	3592
QY	237	pThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValI	257
DB	3591	GACGAATGAAGTACCTTCGACGACAGATCATGGCACCCTTCGTCGCGGGGTAAT	3532
QY	257	eAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgVa	277
DB	3531	CGCTTCTCCAGGGAATGCTAGGCTTCCTCCGACGCCGATCTTTACATATTAAAGT	3472
QY	277	lPheThrAsn	280
DB	3471	TTTTACGAACCTCCCAAGTGAGCAACGATAGAACTTAACTAAGCCAATCTAATTAG	3412
QY	281	-----Asn-GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleL	298
DB	3411	CCCTACATCGGCGAGGTTTCTTACCTTCTGCTTCTGATGCTTCACTCACTACGGATAT	3352
QY	298	eulLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProp	318
DB	3351	ATAGGAAATAACATCTCAACCTTAGCATTTGGGGTCCGACTTTATGGACTCCCGCT	3392
QY	318	heValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyA	338
DB	3291	TCGTTCAAAAGGTTTGGAACTGTCGCCTAAATATGTCAATATGATATGGCAGCAGAA	3332
QY	338	snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV	358
DB	3231	ATGATGCTCCCTTGTATCGGCACGCTAAACAATCTCTGCGCATCAGACGATGTATTTGGCG	3172
QY	358	aIGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr	378
DB	3171	TTGGTGGCATTCAGTTTGTATTAATCGCAAGTTTAGTTCGACAGGATGACACAGT	3112
QY	378	rpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyV	398
DB	3111	GGGAACCTTCCCTTAGGCTACGGACGATGGGACTCGATATTTGTCAGTACGGAAGTCAAG	3052
QY	398	aIArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerP	418
DB	3051	TGAAGGCGAGTGATGTCGCAAGGGTGCAGACGACTCTCTGGAACATCCGTCTCTCTC	2992
QY	418	roValValAlaGlyAlaValThrLeuLeuValSer	437
DB	2991	CAGTTGTTCCAGGGGCTGCTGCACCTCTTATAGCGGTCATTTAGAAAATCGACTACA	2932
QY	437	aIAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyVala	457
DB	2931	TAAACCCAGCATCTCTTAAAGCAGGTACTCATTAAGGTGCGGAGAAATGCCGCAATTATA	2872
QY	457	snMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaThrGlnIleLeuAsnS	477
DB	2871	ACATGTTTACGAGGAGCTGGAATACTGAATTTGCTGAAGATATGCAAGTATTTCTGT	2812
QY	477	eTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM	497
DB	2811	CATACAAACCAAGATAACCTTATTCGCGCATACCTTCACCTCACCCAA--AACTATA	2755
QY	497	eTTrpProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValT	517
DB	2754	TGTGGCTTATAGTCCCAACCTCTGTACTATGGAAGCTCCGTCGCTATTGCAACGTTA	2695
QY	517	hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProT	537
DB	2694	CCATCTCAATGGTATCTCTGTCAAGTCATATAGTTGGCATCCCTCAATGGATTTCCCG	2635
QY	537	yrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpP	557
DB	2634	ATTTGGAACCAAGGTCAGTTCTTCAAGTATTCACCAAGTTTCGCCATCGTTTGGC	2575
QY	557	roTTrpSerGlyTyrLeuAlaIleSerIle	566
DB	2574	CGTGACCGGTTGGATGTCAGTTTATTGTTGGTGAAGTATGCAAAATAGCAACGCTATAA	2515
QY	566	-----	566
DB	2514	AGGATATATGGGTACATATTGATATGAAGTGTAAAGCCATTTCTATATACCAACAATTTT	2455
QY	567	-----ValThrLysLysAlaLeuSerTrpGluGlyIleAlaGlnGlyHisVal	583
DB	2454	TGTTTACCAGCTGTAAGAAAGGAGGAAACTTTGAAGGTTGTTGTAAGAAAGTATC	2395
QY	584	MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer	603
DB	2394	ACCTTAGTTTGGAAAGCTTTAAACAGACACCAACGAAACT-----CATGTTACA	2344
QY	604	ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal	623
DB	2343	GAAGTCGACTTCTTCTTACAAATAAAGGTTACTCCAAAACCCGCAAGAAAGAGGATT	2284


```
QY 624 LeuTyrPaspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAsnLeu 643
|||||
Db 2283 TTATGGGATCAGTACCACAGCCTAAGTATCCACCGCTATATCCACGAGATGATCTC 2224
QY 644 ArgMetLysAsnAspProLeuAspTyrPaspGlnHisIleHisThrAsnPheArgasp 663
|||||
Db 2223 AAAGTTAACTAGATCTCTGGACTGGAGGCGACATATACACAAACTTTAGGGAC 2164
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
|||||
Db 2163 ATGTATACACATTTACGAAATGTTGGCTACTACATTTGTTGGAGAACCTTCACC 2104
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlyTyrPhe 703
|||||
Db 2103 TGCCTCAATGCTCGGATTTAGCGCGTATTGATGTTGACCTGAGAGAGGTTTGGC 2044
QY 704 ProGluGluIleAlaLysLeuArgAspVal---AspAsnGlyLeuSerLeuValIle 722
|||||
Db 2043 GACGAGGAATTAACGCTTTACGGAACAGCTGTATAAAGAGCGTTGAATGTCGGTGA 1984
QY 723 PheSerAspTyrPaspThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
|||||
Db 1983 TTCGAGACTGGTATACACACCTGTGATGAAAAAATTAATTTCTTTGACGAGAACACC 1924
QY 743 ArgGlnTyrPaspMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeu 762
|||||
Db 1923 CGACANTGGTGACACCCGACACTGGTGGCGCAATATTCAGCGCTTGAATGATTATTG 1864
QY 763 SerValTyrAsnMetGlyPheSerAspGlyLeuTyrGlyGluPheThrLeuAlaAsn 782
|||||
Db 1863 AAGCATTTGGAATTTGCTTTGGCGATTTTTCGGTGGAGGACATTTCAAACTGGCGGAC 1804
QY 783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAsp-----Gly 800
|||||
Db 1803 CATTCATGACTATGCTAGTGGAGCCCAATTTGAAGTTTCCAATCAATCCAGGAGAT 1744
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuVal----- 814
|||||
Db 1743 ATTATAGTGGCGCAAACTGAATGACCAAGGACTTTGGTAAATTTAATAATCCTTT 1684
QY 815 -----LeuLysGlnGluThr 819
|||||
Db 1683 TTAGTTTCCAGAAATACAACTTTGCCTTTTCCAGATTTAATCTTAAACACCC-AGC 1625
QY 820 AlaValValGlu---AspValProIleLeuGlyLeuTyrGlnIleProAla----- 835
|||||
Db 1624 AAGGTAGCAAACTAGATCTACCTATTTTGGTGTGTTCCAAACCAAGGCGAACAGTATT 1565
QY 835 ----- 835
Db 1564 CAACGACAGGAGAAATCGTGTCAATGCGGAAAGCAATTTGGCAGAGGCTATACCACA 1505
QY 835 ----- 835
Db 1504 GATTACTCCATTTAAGAACCGGTTTTCCTACTCGCAACGACCAAGCAAGATATCAGT 1445
QY 836 -----GluGlyGlyArgIleValLeuTyrGlyAsp 846
|||||
Db 1444 TTTGCGAAAAGCAATATCATGAACCTAAGATGAAGGACGCTATTGCGGTATATGGGAC 1385
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTyrPheLeuLeuAspAlaLeu 866
|||||
Db 1384 TCCAACTGCTCGACTCCACGATCTGAGAGGCTTGTCTACTGCTGTCTAATAACGTTT 1325
QY 867 LeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGln 886
|||||
Db 1324 TTAGATTTCATAAATAC-----TCGAC----- 1301
QY 887 ArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHis 906
|||||
Db 1300 -----AAATCAAGTTTATTGCGAG 1283
QY 907 ArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
|||||
```

RESULT 10

```
AAAX39701
ID AAX39701 standard; DNA; 786 BP.
XX
AC AAX39701;
XX
DT 02-JUL-1999 (first entry)
XX
DE Gastric cancer associated gene.
XX
KW Cancer associated antigen; diagnosis; research; treatment; human;
KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
KW prostate cancer; ss.
XX
OS Homo sapiens.
XX
PN W09904265-A2.
XX
PD 28-JAN-1999.
XX
PF 15-JUL-1998; 98WO-US14679.
XX
PR 22-JUN-1998; 98US-0102322.
PR 17-JUL-1997; 97US-0896164.
PR 10-OCT-1997; 97US-0061599.
PR 10-OCT-1997; 97US-0061765.
PR 10-OCT-1997; 97US-0948705.
PR 11-OCT-1997; 97GB-0021697.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Chen Y, Gout I, Gure A, O'Hare M, Obata Y, Old LJ;
PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E;
PI Tureci O;
XX
DR WPI; 1999-132448/11.
XX
PT New isolated cancer associated nucleic acids and polypeptides -
PT isolated using sera from cancer patients, used to develop products
PT for the diagnosis, monitoring or treatment of cancers
XX
PS Claim 67; Page 516; 787pp; English.
XX
CC The invention relates to a method for diagnosing a disorder characterised
CC by expression of a human cancer associated antigen precursor coded for by
CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
```


CC biological sample isolated from a subject with an agent that specifically
CC binds to the NAM, an expression product or a fragment of an expression
CC product complexed with an HLA molecule; and (b) determining the
CC interaction between the agent and the NAM or the expression product as a
CC determination of the disorder. The products and methods can be used in
CC the diagnosis, monitoring, research, or treatment of conditions
CC characterised by the expression of various cancer associated antigens.
CC The invention provides nucleic acid sequences and encoded polypeptides
CC which are cancer associated antigen precursors expressed in human breast
CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
CC lung cancer.

XX Sequence 786 BP; 202 A; 190 C; 214 G; 172 T; 8 other;

Alignment Scores:
Pred. No.: 1-26e-77 Length: 786
Score: 1229.50 Matches: 239
Percent Similarity: 91.92% Conservative: 0
Best Local Similarity: 91.92% Mismatches: 18
Query Match: 21.89% Indels: 4
DB: 20 Gaps: 1

US-09-830-837-6 (1-1052) x AAX39701 (1-786)

QY 692 ThrLeuLeuMetValAspSerGluGluGluTyrPheProGluGluLeuAlaLysLeuArg 711
DB 3 ACTTCTGATGTTGACAGTGGAGGAGGAGTACTTCCCTGAAGAGATCGCAAGCTCCGG 62
QY 712 ArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerVal 731
DB 63 AGGAGCTGGACAAACGCCCTCTCGCTCGTCACTTTCAGTCACTGGTACAACTTCTGTT 122
QY 732 MetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGly 751
DB 123 ATGAGAAAGTGAAGTTTATGATGAACACCAAGCAGTGGTGGATCGCGGATACCGGA 182
QY 752 GlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAsp 771
DB 183 GGAGCTAAACATCCAGCTCTGATGAGTCTGCTGTGTGGAACATGGGTTCAGCGAT 242
QY 772 GlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTrpAlaSerGlyCys 791
DB 243 GGCTGTATGAAGGGAGTTCACCTTGGCCCAACCATGACATGATTTATCGCTCAGGGTGC 302
QY 792 SerIleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGly 811
DB 303 AGCATCGGAAGTTTCCAGNAGATGGCTCGTGTATACACAGACTTTCAGAGCAACGA 362
QY 812 LeuGluValLeuLysGlnGluThrAlaValAlaValGluAsnValProIleLeuGlyLeuTyr 831
DB 363 TTGGAGGTTTAAAGCAGGAACAGCAGTGTGTTGAAACGCTCCCATTTTGGGACTTAT 422
QY 832 GlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAsp 851
DB 423 CAGATTCACAGTCCAGGGTGGAGCCGATGTGATCTATGGGACATCCCAATTCCTGGAT 482
QY 852 AspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuGlnTyrThrSer 871
DB 483 GACAGTCACGACAGAGGAGTCTTTTGGCTTCTGGATGCCCTCTCCAGTACACATCG 542
QY 872 TyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGlnArgPro-ProSerGln 891
DB 543 TATGGGTGACACCGCTAGCTCAGTCACTCTGGGAACGCCAGCGCCCTCCANTTGG 602
QY 891 yAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysVa 911
DB 603 AGCAAGCTCAGTCCATCCAGAGAGGATGGAAGAAACCATCT-CATCGGTACTCTCAAGGT 661
QY 911 lLeuGluAlaHisAlaGlyAspProLysProArgProLeuProAlaCysProArgLeuSe 931
DB 662 TCTGGAGGCCATTTGGGAACCAACCAACCTCGGGCTCNACACCTGTCCANGCCTGTN 721
QY 931 rTTPAlaLysPro-GlnProLeuAsnGluThr---AlaProSerAsnLeuTrpLys 948

Db 722 CTGGGCCAAGCCCAAGACCTTTAAACCAANAACGCGCCCAATTAACCTTTGGAAA 777
RESULT 11
AAZ15839
ID AAZ15839 standard; cDNA; 757 BP.
XX AAZ15839;
AC AAZ15839;
XX 12-OCT-1999 (first entry)
DT Human gene expression product cDNA sequence SEQ ID NO:3308.
XX Human; gene; gene expression product; diagnosis; therapy; probe;
DE detection; mapping; tissue typing; profiling; forensic; cancer;
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.
XX Homo sapiens.
XX WO9938972-A2.
PN 05-AUG-1999.
PD 28-JAN-1999; 99WO-US01619.
PF 03-APR-1998; 98US-0080666.
PR 28-JAN-1998; 98US-0072910.
PR 24-FEB-1998; 98US-0075954.
PR 31-MAR-1998; 98US-0080114.
PR 03-APR-1998; 98US-0080515.
XX (CHIR) CHIRON CORP.
PA (HYSE-) HYSQ INC.
PA Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;
XX WPI; 1999-494092/41.

XX Novel human genes and their expression products which are
PT differentially expressed in different cell types
PS Claim 1; Page 1586; 2479pp; English.
XX The present invention describes a library of human polynucleotides
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is
CC a method of detecting differentially expressed genes correlated with the
CC cancerous state of a mammalian cell, comprising detecting at least one
CC differentially expressed gene product in a test sample from a cell
CC suspected of being cancerous, where the gene product is encoded by one
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The
CC polynucleotides can be used as a source of primers and probes, which can
CC be used for a variety of purpose, e.g. detection of expression levels,
CC mapping, tissue typing or profiling, forensics, genetic analysis and
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides
CC can be used for raising antibodies for experimental, diagnostic and
CC therapeutic purposes. The polynucleotides may also be used to construct
CC arrays for diagnostics (which may be used to determine function of an
CC encoded protein); and to detect differences in expression levels between
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to
CC identify a genetic predisposition or susceptibility to a disease such as
CC cancer). The polynucleotides of the invention are especially used in the
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,
CC and lung cancer. The polynucleotides can also be used to screen for
CC peptide analogues and antagonists.
XX Sequence 757 BP; 157 A; 198 C; 187 G; 161 T; 54 other;
SQ Alignment Scores:
Pred. No.: 2.09e-72 Length: 757

RESULT 13
AAI99218
ID AAI99218 standard; DNA; 9689 BP.
XX
XX
AC AAI99218;
XX
DT 07-JAN-2002 (first entry)
XX
XX Human excretory related polynucleotide SEQ ID NO 982.
XX
KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine;
KW excretory system; ds.
OS Homo sapiens.
XX
XX WO200155313-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01323.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180528.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX 19-MAY-2000; 2000US-0205515.
XX 07-JUN-2000; 2000US-0209467.
XX 28-JUN-2000; 2000US-0214886.
XX 30-JUN-2000; 2000US-0215135.
XX 07-JUL-2000; 2000US-0216647.
XX 07-JUL-2000; 2000US-0216880.
XX 11-JUL-2000; 2000US-0217487.
XX 11-JUL-2000; 2000US-0217496.
XX 14-JUL-2000; 2000US-0218290.
XX 26-JUL-2000; 2000US-0220963.
XX 26-JUL-2000; 2000US-0220964.
XX 14-AUG-2000; 2000US-0224518.
XX 14-AUG-2000; 2000US-0224519.
XX 14-AUG-2000; 2000US-0225213.
XX 14-AUG-2000; 2000US-0225214.
XX 14-AUG-2000; 2000US-0225266.
XX 14-AUG-2000; 2000US-0225267.
XX 14-AUG-2000; 2000US-0225268.
XX 14-AUG-2000; 2000US-0225270.
XX 14-AUG-2000; 2000US-0225447.
XX 14-AUG-2000; 2000US-0225757.
XX 14-AUG-2000; 2000US-0225758.
XX 14-AUG-2000; 2000US-0225759.
XX 18-AUG-2000; 2000US-0226279.
XX 22-AUG-2000; 2000US-0226681.
XX 22-AUG-2000; 2000US-0226868.
XX 22-AUG-2000; 2000US-0227182.
XX 23-AUG-2000; 2000US-0227009.
XX 30-AUG-2000; 2000US-0228924.
XX 01-SEP-2000; 2000US-0229287.
XX 01-SEP-2000; 2000US-0229343.
XX 01-SEP-2000; 2000US-0229344.
XX 01-SEP-2000; 2000US-0229345.
XX 05-SEP-2000; 2000US-0229509.
XX 05-SEP-2000; 2000US-0229513.
XX 06-SEP-2000; 2000US-0230437.
XX 06-SEP-2000; 2000US-0230438.
XX 08-SEP-2000; 2000US-0231242.
XX 08-SEP-2000; 2000US-0231243.
08-SEP-2000; 2000US-0231244.
08-SEP-2000; 2000US-0231413.
08-SEP-2000; 2000US-0231414.
08-SEP-2000; 2000US-0232080.
08-SEP-2000; 2000US-0232081.
12-SEP-2000; 2000US-0231968.
14-SEP-2000; 2000US-0232397.
14-SEP-2000; 2000US-0232398.
14-SEP-2000; 2000US-0232399.
14-SEP-2000; 2000US-0232400.
14-SEP-2000; 2000US-0232401.
14-SEP-2000; 2000US-0233063.
14-SEP-2000; 2000US-0233064.
14-SEP-2000; 2000US-0233065.
21-SEP-2000; 2000US-0234223.
21-SEP-2000; 2000US-0234274.
25-SEP-2000; 2000US-0234997.
25-SEP-2000; 2000US-0234998.
26-SEP-2000; 2000US-0235484.
27-SEP-2000; 2000US-0235834.
27-SEP-2000; 2000US-0235836.
29-SEP-2000; 2000US-0236327.
29-SEP-2000; 2000US-0236367.
29-SEP-2000; 2000US-0236368.
29-SEP-2000; 2000US-0236369.
29-SEP-2000; 2000US-0236370.
02-OCT-2000; 2000US-0236802.
02-OCT-2000; 2000US-0237037.
02-OCT-2000; 2000US-0237038.
02-OCT-2000; 2000US-0237039.
13-OCT-2000; 2000US-0237040.
13-OCT-2000; 2000US-0239935.
13-OCT-2000; 2000US-0239937.
20-OCT-2000; 2000US-0240960.
20-OCT-2000; 2000US-0241221.
20-OCT-2000; 2000US-0241785.
20-OCT-2000; 2000US-0241786.
20-OCT-2000; 2000US-0241787.
20-OCT-2000; 2000US-0241808.
20-OCT-2000; 2000US-0241809.
20-OCT-2000; 2000US-0241826.
01-NOV-2000; 2000US-0244617.
08-NOV-2000; 2000US-0246474.
08-NOV-2000; 2000US-0246475.
08-NOV-2000; 2000US-0246476.
08-NOV-2000; 2000US-0246477.
08-NOV-2000; 2000US-0246478.
08-NOV-2000; 2000US-0246523.
08-NOV-2000; 2000US-0246524.
08-NOV-2000; 2000US-0246525.
08-NOV-2000; 2000US-0246526.
08-NOV-2000; 2000US-0246527.
08-NOV-2000; 2000US-0246528.
08-NOV-2000; 2000US-0246532.
08-NOV-2000; 2000US-0246609.
08-NOV-2000; 2000US-0246610.
08-NOV-2000; 2000US-0246611.
08-NOV-2000; 2000US-0246613.
17-NOV-2000; 2000US-0249207.
17-NOV-2000; 2000US-0249208.
17-NOV-2000; 2000US-0249209.
17-NOV-2000; 2000US-0249210.
17-NOV-2000; 2000US-0249211.
17-NOV-2000; 2000US-0249212.
17-NOV-2000; 2000US-0249213.
17-NOV-2000; 2000US-0249214.
17-NOV-2000; 2000US-0249215.
17-NOV-2000; 2000US-0249216.
17-NOV-2000; 2000US-0249217.
17-NOV-2000; 2000US-0249218.
17-NOV-2000; 2000US-0249244.
17-NOV-2000; 2000US-0249245.
17-NOV-2000; 2000US-0249264.

Percent Similarity:	41.96%	Conservative:	68
Best Local Similarity:	27.17%	Mismatches:	164
Query Match:	6.07%	Indels:	103
DB:	9	Gaps:	19
US-09-830-837-6 (1-1052) x AAN80745 (1-1220)			
QY	48	SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys	67
DB	97	AGCAGTACA	132
QY	68	AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle	87
DB	133	CAGACATGAGTGCATGAGTTCGCCAAGAAAAGGATGTTAT	177
QY	88	IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys	107
DB	177		177
QY	108	GluLysGlnLysAlaGlyLeuLeuThrLeuGlu	126
DB	178	---TCTGAAAAGCGGAGAGTTCAAAACCAATTTAAGTATGTTAACGGCGCGCAGCA	234
QY	127	ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro	146
DB	235	ACATTGGATGAAAAGCTGTAAAGAATTGAAA	285
QY	147	CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSer	166
DB	285		285
QY	167	LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu	186
DB	286	---TAATGGAAGAGATCATATTCATGCAATGATATGCG	321
QY	187	ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly	206
DB	322	CAATCTGTTCTTATGCAATTTCTCAA	378
QY	207	TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro	226
DB	379	TACACAGGCTCAACGTAAAGTAGCTGTATGACACAGCGGAATGACTCTTCTCAACCT	438
QY	227	HisPheLysAsnValLysGluArgTrpAsnTrpThrAsnGluArgThr	244
DB	439	GACTTA	495
QY	245	AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu	262
DB	496	GACGGCAGTTCTCACGGTAGCGATGACGGGTACGATTCGGCTCTTAATAACTCAATC	555
QY	263	---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn	281
DB	556	GGTGTCTGGCGTACGCCACCGCATCATATATGCAAGTAAAGTGTGATTAACA	615
QY	282	GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle	301
DB	616	GGAAGCGCAATATAGCTGGATATTACGGCATTGAGTGGCCATTTCACACATATG	675
QY	302	AspValLeuAsnLeuSerIleGlyProAspPheMetAsp	319
DB	676	GATGTTATCAATCATGAGCTTGGCGGACCTTCTACAGCGCTGAAACAGTGGTT	735
QY	320	AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp	339
DB	736	GACAAAGGCC	786
QY	340	GlyPro	357
DB	787	GGTTCATCCGAGCACACACAGTCGGTACCTGCAAAATATCCTTCTACTATTGCA	846
QY	358	ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrThr	377
DB			
DB	847	GTAGTGGCGGTAAACAGCAGCAACAAAGAGCTTCATTCTCCAGCGCAGGTTCT	900
QY	378	TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly	397
DB	901	---GAGCTT	930
QY	398	ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer	417
DB	931	ATCCAAAGCACCTCTCGGAGCAGCTTACGGCGCTTATACGGAACGTCATCGGCACT	990
QY	418	ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal	437
DB	991	CCTCACGTTGCGGAGCAGCAGCGTTAATCTTTCT	1041
QY	438	AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn	457
DB	1042	ACAAACGCGAAGTCCGTGATGTTAGAAAGCACTGCACATATCTTGA	1095
QY	458	MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe	477
DB	1096	TCITTTCTACTATGAAAGGGTTAATCAACGTAACGACGCTGCACAT	1149
QY	477	rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyr	496
DB	1150	TAAAGAAGCAGGTT	1198
RESULT 16			
AAQ06587	standard; DNA; 1220 BP.		
AC	AAQ06587;		
DT	18-FEB-1991 (first entry)		
DE	apra gene encoding subtilin.		
KW	Detergents; depilatory tanning; serology; ds.		
OS	Bacillus subtilis.		
FT	Key	Location/Qualifiers	
FT	conflict	568..573	
FT		/tag= a	
FT		/label= gtagc	
FT		/note= "Stahl et al. J. Bacteriol., 158, 411-418 (1984)."	
FT	CDS	1..1143	
FT		/tag= b	
FT	mat_peptide	319..1143	
FT		/tag= c	
PN	EP398539-A.		
XX	22-NOV-1990.		
XX	01-MAY-1990;	90EP-0304715.	
XX	17-MAY-1989;	89US-0353124.	
XX	(AMGE-) AMGEN INC.		
XX	Zukowski MM, Narhi LO, Levitt M;		
XX	WPI; 1990-350298/47.		
XX	P-PSDB; AAR07970.		
XX	Bacillus subtilin analogues - with improved pH thermal and oxidn.		
XX	stability useful in cleaning compns.		
XX	Claim 20; Table 1; 39pp; English.		
XX	Modified analogues of subtilisin are useful in cleaning fabrics, and		
XX	have an improved resistance to oxidation, heat and pH extremes.		
XX	Analogues have one or more negative AAs present in the calcium		

DR P-PSDB; AAF60571.

XX Bacillus strains with reduced extra-cellular protease levels -
 PT useful as hosts for secretion of heterologous polypeptide(s) and
 PT proteins.

XX Example; Fig 2; 30pp; English.

XX In the example, the inventors inactivate the apr gene by insertion
 CC of a functional cat gene coding for chloramphenicol
 CC acetyltransferase from *S. aureus*. The inserted cat gene confers
 CC chloramphenicol resistance, thus facilitating the selection of
 CC transformants. The reduced levels of subtilisin greatly reduces the
 CC likelihood that a secreted foreign protein will be proteolytically
 CC degraded prior to recovery.

XX Sequence 1524 BP; 461 A; 329 C; 331 G; 403 T; 0 other;

Alignment Scores:

Pred. No.:	9,46e-15	Length:	1524
Score:	341.00	Matches:	125
Percent Similarity:	41.96%	Conservative:	68
Best Local Similarity:	27.17%	Mismatches:	184
Query Match:	6.07%	Indels:	103
DB:	7	Gaps:	19

US-09-830-837-6 (1-1052) x AAN60475 (1-1524)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 DB 379 AGCATACAC-----GAAAGAAATACATGTCGGATT-----AAA 414
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnIrpArgIle 87
 DB 415 CAGACAAATGAGTCCCATGATTCGCCCAAGAAAGGATGTTAT----- 459
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 DB 459 ----- 459
 QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu-----AspHisProAsnIleLysArgVal 126
 DB 460 ---TCTGAAAAGCGGGAAGGTTCAAAAGCAATTTAAGTATGTTAAACGCCGCCGACGA 516
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 DB 517 ACATTGGATGAAGAAGCTTAAGAAGATTCGAA-----AAGATCCGAGCGTTGCA 567
 QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
 DB 567 ----- 567
 QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
 DB 568 -----TATGTGGAAGAGATCATATTGCACATGAATATGCG 603
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 DB 604 CAATCTGTCTTATGGCAATTCCTCA-----ATAAAGCGCCGCTCTTCACCTCAAGC 560
 QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 DB 661 TACACAGGCTCTAACGTAAGTAGCTGTATCGACAGCGGAATTGACTCTCTCATCCT 720
 QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 DB 721 GACTTA---AGCTCAGAGCGGAGCAGACGCTTCGTACCTTCCTCAACCAACCAATACACAG 777
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 DB 778 GACGGCAGTCTCAGCGGATGCGATGAGCGCGATGAGATTCGCTCTTAATTAATCAATC 837
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281

DB 838 GGTGTTCTGGCGTAGCGCAAGCGCATATTATATGTCAGTAAAGTCTGTGATTCACA 897
 QY 282 GlnValSerTyrThrSerIrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
 DB 898 GGAAGCGGCAATATAGCTGGATTATTAACGCGATTGAGTGGCCATTCCACAAATATG 957
 QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
 DB 958 GATGTATACATAGACCTTGGCGACCTACTGTTCTACAGCGCTGAACACAGTCGTT 1017
 QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
 DB 1018 GACAAAGCC-----GTTTCCAGCGGTATGCTGCTGCTCCGACGCGGAACGAA 1068
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspClnMetAspValIleGly 357
 DB 1069 GGTTCATCCGGAAGCACAGCACAGTCGCTACCTGCAAAATATCTCTACTATTCGA 1128
 QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
 DB 1129 GTAGTGGCGTAAACAGACCAACAGAGCTTCATTTCCAGCCGAGTTCT----- 1182
 QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
 DB 1183 ---GAGCTT-----GATGTGATGCTCTCTGCGGTGTC 1212
 QY 398 ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
 DB 1213 ATCCAAAGCACACTTCCTGGAGCGCTTACGGCGCTTATAACGGAACGTCCTGCGGACT 1372
 QY 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
 DB 1273 CTTCACTGTCGGGAGCAGCAGCGTTAATCTTCT-----AAGCACCGGACTGG 1323
 QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
 DB 1324 ACAAGCGCGCAAGTCGCTGATGCTTTAGAAAGCACTGCAACATATCTTGA-----AAC 1377
 QY 458 MetPheGluGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
 DB 1378 TCTTTCTACTATGGAAGGGTTAATCAACGTACAAGCGCTGCACAAAT-----AATAG 1431
 QY 477 rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyr 496
 DB 1432 TAAAGAGACGAGCTT-----CCTCCATACCTGCTCTTTTATTGTCAGCAT 1480
 RESULT 18
 AAN81505
 ID AAN81505 standard; DNA; 2044 BP.
 XX
 AC AAN81505;
 DX 10-SEP-1990 (first entry)
 XX
 DE DNA encoding the alkaline protease gene promoter domain, and domain
 DE relating to protein secretion of *Bacillus licheniformis*.
 XX
 KW Alkaline protease gene promoter domain; *Bacillus licheniformis*;
 KW *Bacillus subtilis* PW10; alkaline protease purification.
 XX
 OS *Bacillus licheniformis*.
 FH Key Location/Qualifiers
 FT Promoter 149..150
 FT /*tag= a 151..154
 FT /*tag= b 172..177
 FT /*tag= c 209..213
 FT /*tag= d 1389..1400
 FT repeat_unit

CC subtilisins can be used as the second part of the fusion protein. The
 CC mutated versions of subtilisin are incapable of autolytic
 CC maturation. The DNA fusion sequences are useful for the production of
 CC heterologous proteins. As the subtilisin mutants are incapable of
 CC autolytic maturation, they are therefore bound to the Bacillus cell
 CC membrane. The inability of the mutant protein to leave the cell membrane
 CC means that they are not able to degrade desired proteins when
 CC heterologous protein production is being carried out in Bacillus cells.
 CC Heterologous proteins are sequestered at cell membranes and can be easily
 CC isolated and released by the action of enzymatically active subtilisin.
 XX
 SQ Sequence 1500 BP; 443 A; 343 C; 328 G; 386 T; 0 other;

Alignment Scores:

Pred. No.: 1.39e-14 Length: 1500
 Score: 338.50 Matches: 133
 Percent Similarity: 40.58% Conservative: 76
 Best Local Similarity: 25.83% Mismatches: 180
 Query Match: 6.03% Indels: 126
 DB: 20 Gaps: 21

US-09-830-837-6 (1-1052) x AA206712 (1-1500)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 DB 233 AGCGTACA-----GAAAGAAATACATTCTCGATT-----AAA 268
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
 DB 269 CAGACAAATGAGTGCATGAGTTCCCGCCAAAGAAAGATGTTATT----- 313
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 DB 313 ----- 313
 QY 108 GluLysGlnLysAlaGluLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 DB 314 ---TCTGAAAAGCGGGAAGGTTCAAAAGCAATTAAGTATGTTAAGCGCGCGCAGCA 370
 QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 DB 371 ACATTGGATGAAAAGCTGTAAGAAGATTGAAA-----AAGATCCGAGCGTTGCA 421
 QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
 DB 421 ----- 421
 QY 167 LeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
 DB 422 -----TATGTGGAAGAGATCATATTGCACATGATATGCG 457
 QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 DB 458 CAATCTGTCTTATGCGCATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGGC 514
 QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 DB 515 TACACAGGCTTAACGTAAAGTAGCTGTTATCGACAGCGGAATTGACTCTTCTCATCT 574
 QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 DB 575 GACTTA---AACGTACAGCGCGGAGCAAGCTTCGTACCTTCTGAAACAAACCCATACCAG 631
 QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 DB 632 GACGGCAGTTCTCAGGTAGCATGTAGCGGTACGATTGCGGCTCTTAACTCAATC 691
 QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
 DB 692 GTGTCTTGGCGGTAGCCCAAGCCATCATATATGCGAGTAAAGTCTGATTCAACA 751
 QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
 DB 301 ----- 301

DB 752 GGAAGCGGCCAATATACCTGGATTATTAAACGGCATTTGAGTGGCCATTTCCAAACAATATG 811
 QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMetAsp-----HisProPheVal 319
 DB 812 GATGTATCAACATGAGCTTGGCGAAGCTACTGTTCTACAGCGTGAACAGTCTGCT 871
 QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
 DB 872 GACAAAGCC-----GTTTCCAGCGTATCGTCTGCTGCCGACGCGGAACGAA 922
 QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
 DB 923 GGTTCATCGGAGCAACAGCAGTACCTGCTCAAAATATCTCTTCTATTATTCGA 982
 QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
 DB 983 GTAGTGGCGGTAAACAGCAGCAACAAAGAGCTTCATCTCCAGCGAGTCT----- 1036
 QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTrpGlyAlaGly 397
 DB 1037 ---GAGCTT-----GATGATGATGCTCTCGCGTGTC 1066
 QY 398 ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerSerArgGlyThrSerValAlaSer 417
 DB 1067 ATCCAAAGCAGCTTCTGAGGAGCTTACGCGCTTATACCGAAGCTTCCAGCGACT 1126
 QY 418 ProValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
 DB 1127 CTTCACTGCTGCGGAGCAGCGTAAATCTTCT-----AAGCAGCCGACTTGG 1177
 QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsn 457
 DB 1178 ACAACGCGCAAGTCCGTGATGCTTTAGAAAGCAGCTCAACATATCTTGA-----AAC 1231
 QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAla---TyrGlnIleLeuAsnSe 477
 DB 1232 TCTTCTACTATGGAAGGTTAATCAACGCTACAAGCAGCTGCACAAAT-----AATAG 1285
 QY 477 rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMe 497
 DB 1286 TAAAGAAGCAGGCT-----CCTCCATACCTGCTCTCTTTTATTGTGACAT-- 1334
 QY 497 ttrpTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnValTh 517
 DB 1335 -----CCTGATGTTCCGCGCATTC 1354
 QY 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgIle 528
 DB 1355 TCTTCTTCTCCGATGTTGAATCCGTCATGATCGAGGATGGTGGCTCTGAAATC 1414
 QY 528 eValAspLysProAspTrpGlnProTyrLeuProGlnAsnGly 542
 DB 1415 TTCACAGCAGCGGAGGATCAACCTGCTCAGCGCGCGTACGCGC 1457
 RESULT 22
 AAQ03536
 ID AAQ03536 standard; DNA; 1499 BP.
 XX
 AC AAQ03536;
 XX 16-AUG-1990 (first entry)
 DT
 DE Subtilisin gene.
 XX
 KW Carbonyl hydrolase; subtilisin; neutral protease; ds.
 XX
 OS Bacillus subtilis.
 XX
 FH Key Location/Qualifiers
 FT CDS 137..1280
 FT /*tag= a
 PN EP357157-A.

XX 07-MAR-1990.
XX PD
XX PF
XX PR 22-JUN-1984; 84EP-0202584.
XX PR 24-JUN-1983; 83US-0507419.
XX PR 29-MAY-1984; 84US-0614612.
XX PR 29-MAY-1984; 84US-0614615.
XX PR 29-MAY-1984; 84US-0614616.
XX PR 29-MAY-1984; 84US-0614617.
XX PR 29-MAY-1984; 84US-0614491.
XX PA (GETH) GENENTECH INC.
XX PI Bott RR, Estell DA, Ferrari E, Henner DJ, Wells JA;
XX DR WPI; 1990-068909/10.
XX DR P-PSDB; AAR03737.
XX PT Mutant prokaryotic carbonyl hydrolase enzymes -
PT obtd. by site-directed oligo-nucleotide mutagenesis, used in
PT food processing and cleaning industries.
XX PS Claim 16; Fig 7; 39pp; English.
XX CC Probe derived from subtilisin gene was used to. Isolate carboxyl
CC hydrolase gene, mutant versions of which exhibit different oxidative
CC stability and/or pH activity.
XX SQ Sequence 1499 BP; 443 A; 342 C; 328 G; 386 T; 0 other;

Alignment Scores:
Pred. No.: 1-51e-14 Length: 1499
Score: 338.00 Matches: 124
Percent Similarity: 41.96% Conservative: 69
Best Local Similarity: 26.96% Mismatches: 164
Query Match: 6.02% Indels: 103
DB: 11 Gaps: 19

US-09-830-837-6 (1-1052) x AAQ03536 (1-1499)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
DB 233 AGCAGTACA-----GAAAAGAAATACATGTCGGATTT-----AAA 268
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
DB 269 CAGACAAATGAGTGCATGATGTCGCCAAGAAAGAGATGTTATT----- 313
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
DB 313 ----- 313
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
DB 314 ---TCTGAAAGCGCGGAAGGTTCAAAAGCAATTTAAGTATGTTAACGCGCGCAGCA 370
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
DB 371 ACATGGATGAAAGAGCTTAAGAAATGAA-----AAGATCCGAGCGTTCGA 421
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSer 166
DB 421 ----- 421
QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeu 186
DB 422 -----TATGTGGAAGAGATCATATTGCACATGAATATGCG 457
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
DB 458 CAATCGTTCCTATGGCATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGCG 514

QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
DB 515 TACACAGGCTCTAACGTAAAGTAGCTTTATCGACAGCGGAATGACTCTTCTCATCCT 574
QY 227 HisPheLysAsnValLysGluArgTrpAsnTrpAsnGluArgThr-----LeuAsp 244
DB 575 GACTTA---AACGTGAGAGCGGAGCAAGCTTCGTACCTTCTGAAACAAACCCATACCAG 631
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
DB 632 GACGGCAGTCTCAGCGTACGCATGTAGCGGTACGATTCGCCCTCTTAATACTCAATC 691
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
DB 692 GGTGTTCTGGCGCTTAGCCCAAGCGCATCATTTATATGCAAGGCTTGATTCACA 751
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
DB 752 GGAAGCGGCAATATAGCTGGATTATTAACGGCATTGAGTGGGCCATTTCCAAACAATATG 811
QY 302 AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal 319
DB 812 GATGTTATCAACATGAGCTTGGCGACCTACTGTTCTACAGCGCTGAAACAGTCGTT 871
QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
DB 872 GACAAAGCC-----GTTCCAGCGGTATCGTCTGTCGCGCAGCGCGAAGAA 922
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
DB 923 GGTTCATCCGGAAGCACAGCAGTCGGCTACCTCGCAAAATATATCTTCTACTATTGCA 982
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
DB 983 GTAGTGGCGTAAACAGCAGCAACCAAGAGCTTCATCTCCAGCGCAGGTTCT----- 1036
QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
DB 1037 ---GAGCTT-----GATGTGATGGCTCCTCGCGTGTCC 1066
QY 398 ValArgLysSerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
DB 1067 ATCCAAGCACACTTCTCGAGCAGCTTACGGCGCTTATAACGGAAGCTCATCGCGACT 1126
QY 418 ProValValAlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuVal 437
DB 1127 CCTCAGTTCGCGAGCAGCAGCGTTAATCTTTCT-----AAGCACCAGCTGG 1177
QY 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsn 457
DB 1178 ACAACGCGCAAGTCGCTGATCGTTTAGAAAGCACTGCAACATATCTTGA-----AAC 1231
QY 458 MetPheGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
DB 1232 TCTTTCTACTATGGAAGGGTTAATCAACGTACAAGCAGCTGCACAA-----AATAG 1285
QY 477 rTyrLysProGlnAlaSerLeuSerProSerTrpIleAspLeuThrGluCysProTyr 496
DB 1286 TAAAGAGAGCAGGTT-----CCTCCATACCTGCTCTTTTATTGTGTCAGCAT 1334

RESULT 23
AAAX25097/C
ID AAX25097 standard; DNA; 2760 BP.
XX AC AAX25097;
XX DT 05-JUL-1999 (first entry)
XX DE Bacillus subtilis metalloprotease YHFN DNA.
XX KW Metalloprotease; protease; YHFN; textile; detergent; feedstuff;
XX animal feed; host expression system; ss.
XX

OS Bacillus subtilis.
 XX Key Location/Qualifiers
 FH CDS 326..1606
 FT /*tag= a
 XX
 XX WO9914341-A2.
 XX
 XX 25-MAR-1999.
 XX
 XX 08-SEP-1998; 98WO-US18677.
 XX
 XX 15-SEP-1997; 97GB-0019637.
 XX
 XX (GENM) GENENCOR INT BV.
 XX (GENM) GENENCOR INT INC.
 XX
 XX Estell DA;
 XX
 XX WPI: 1999-244036/20.
 XX
 XX P-PSDB; AAY05501.
 XX
 XX New metalloproteases derived from gram-positive microorganisms
 XX
 XX Claim 3; Fig 1A-E; 54pp; English.
 XX
 XX This nucleic acid sequence includes a coding region for a novel
 CC metalloprotease (MP), designated rHN (see AAY05501), of Bacillus
 CC subtilis strain 1-108. rHN has overall amino acid homology to
 CC saccharomyces cerevisiae STE24, and its gene is found immediately
 CC downstream of the major alkaline protease gene (aprE) putative
 CC transcriptional terminator, suggesting a role in regulating and/or
 CC processing the major alkaline protease. The invention relates to
 CC the identification of novel MPs in gram-positive microorganisms,
 CC especially in Bacillus subtilis, Bacillus stearothermophilus,
 CC Bacillus licheniformis and Bacillus amyloliquefaciens. The
 CC invention also provides methods for the production of MP in host
 CC cells as well as for the production of heterologous proteins in a
 CC host cell having a mutation or deletion of part or all of the MP.
 CC Preferred heterologous proteins are hormones, enzymes, growth
 CC factors and cytokines. Gram positive MPs have applications in the
 CC textile industry, in cleaning compositions (claimed), and in animal
 CC feed. They can be used in bar or liquid soap applications, dish
 CC care products, contact lens cleaning solutions and products,
 CC peptide hydrolysis, and waste treatment. MP nucleic acid sequences
 CC can be used to derive probes and primers for the detection of MP.
 XX
 XX SQ Sequence 2760 BP; 680 A; 580 C; 666 G; 834 T; 0 other;
 Alignment Scores:
 Pred No.: 3, 28e-14 Length: 2760
 Score: 338.00 Matches: 124
 Percent Similarity: 41.96% Conservative: 69
 Best Local Similarity: 26.96% Mismatches: 164
 Query Match: 6.02% Indels: 103
 DB: 20 Gaps: 19
 US-09-830-837-6 (1-1052) x AAX25097 (1-2760)
 QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
 DB 2694 AGCAGTACA-----GAAAAGAAATACATTGTCGGATT-----AAA 2659
 QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaAspAsnTrpArgile 87
 DB 2658 CAGACAATGAGTCCCATGAGTCCGCCAAGAAAGATGTTAT----- 2614
 QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
 DB 2614 ----- 2614
 QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
 DB 1695 TCTTCTACTATGAAAGGGTTATCAACGTACACAGCAGCTGCACAAAT-----ATAG 1642

2613 ---TCTGAAAGCGCGAAGGTTCAAAGCAATTTAAGTATGTTAAACCGCGCGACCA 2557
 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
 2556 ACATTGGATGAAAGCTGTAAGCAATGAAA-----AAAGATCCGAGCGTTGCA 2506
 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
 2506 ----- 2506
 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeu 186
 2505 -----TATGTGGAAGAAGATCATATGCAATGATATGCG 2470
 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
 2469 CAATCTGTTCTTATGTCATTTCTCAA---ATTAAAGCGCGGCTTCTTCACTCAAGGC 2413
 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
 2412 TACACAGCTCTAAAGTAAAGTAGCTGTATCGACAGCGAATTGACTTCTCATCCT 2353
 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
 2352 GACTTA---AACGTACAGCGGACGACGCTTCTGTAACAAACCCATACCAG 2296
 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
 2295 GACGCGAGTCTCAGGTACGATGTAGCGGTAGATTGCGCTCTTAATACTCAATC 2236
 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
 2235 GGTGTCCTGGCGGTAGCCCAAGCGCATATATATGATGCAAGAGTGTGATTCAACA 2176
 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
 2175 GGAACGCGCCAAATATAGCTGGATTATTAAACGGCATTAGTGCGCCATTCCCAACAAATG 2116
 302 AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal 319
 2115 GATGTATCAACATGAGCGCTTGGCGACCTACTGGTCTTACAGCGCTGAAACAGTGGT 2056
 320 AspLysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAsp 339
 2055 GACAAGCC-----GTTTCCAGCGGTATCGTCTGTCGCGACGCGAAGCA 2005
 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
 2004 GCTTCATCCGGAACACACACAGCTGCGCTACCTGCAAAATATCTTCTACTATTGCA 1945
 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThr 377
 1944 GTAGTGGGTAAACACAGCAACCAAGAGCTTCATCTCCAGCGAGTTCT----- 1891
 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
 1890 ---GAGCTT-----GATGTGATGGCTCTCGCGCTGCC 1861
 398 ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
 1860 ATCCAAGACACACTTCTCGAGGCGACTTACGCGCTTATAACGGAACGCTCCATCGGACT 1801
 418 ProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
 1800 CCTCAGCTTCCGAGCAGCAGCGTTAATTTCTTCT-----AAGCACCAGCATGG 1750
 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuArgLeuProGlyValAsn 457
 1749 ACAACGCGCACTCGCTGATCGTTTAGAAGACATGCAACATATCTTGA-----AAC 1696
 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuArgAla-TyrGlnIleLeuAsnSe 477
 1695 TCTTCTACTATGAAAGGGTTATCAACGTACACAGCAGCTGCACAAAT-----ATAG 1642

QY 477 rTyrLysProGlnAlaSerLeuSerProSertYrilleAspLeuThrGluCysProTyr 496
 Db 1641 TAAAGAAGCAGGTT-----CCTCATACCTGCTCTTTTATTGTCAGCAT 1593
 RESULT 24
 AAD19518
 ID AAD19518 standard; DNA; 1180 BP.
 XX AC AAD19518;
 XX DT 18-DEC-2001 (first entry)
 XX DE Subtilisin E DNA, used to generate single-stranded DNA template.
 XX KW Template-mediated recombination; nucleic acid isolation; subtilisin E;
 XX KW therapeutic protein; vaccine; ds.
 XX OS Unidentified.
 XX FH Key
 XX FT misc_feature 1..16
 XX FT /tag= a
 XX FT /note= "Sequences that optionally do not correspond
 XX FT to wild-type sequences which may be replaced by upstream
 XX FT regulatory regions and vector supplied sequences
 XX FT depending on the cloning system in use"
 XX FT misc_feature 1163..1180
 XX FT /tag= b
 XX FT /note= "Sequences that optionally do not correspond
 XX FT to wild-type sequences which may be replaced by upstream
 XX FT regulatory regions and vector supplied sequences
 XX FT depending on the cloning system in use"
 XX PN WO200164864-A2.
 XX PD 07-SEP-2001.
 XX PF 28-FEB-2001; 2001WO-US06775.
 XX PR 28-FEB-2000; 2000US-185244P.
 XX PR 29-FEB-2000; 2000US-185815P.
 XX PR 01-MAR-2000; 2000US-186247P.
 XX PR 02-MAR-2000; 2000US-186482P.
 XX PR 06-SEP-2000; 2000US-0656549.
 XX PA (MAXY-) MAXYGEN INC.
 XX PI Affholter JA, Cox A, Ness JE, Carr B;
 XX WP: 2001-616208/71.
 XX DR Generating chimeric nucleic acids to produce therapeutics comprises
 XX PT hybridizing nucleic acids and nicking and elongating regions that are
 XX PT non-hybridized -
 XX PS Disclosure; Fig 6; 172pp; English.
 XX CC The invention relates to a method of single-stranded nucleic acid
 XX CC template-mediated recombination and nucleic acid fragment isolation.
 XX CC Methods include polymerase and polymerase-free recombination of
 XX CC nucleic acid fragments to generate chimeric nucleic acid sequences.
 XX CC This method is useful for isolating nucleic acid fragments,
 XX CC combinatorially assembling nucleic acids and for producing chimeric
 XX CC nucleic acids. The method is useful for single stranded nucleic
 XX CC acid template mediated recombination and nucleic acid fragment
 XX CC isolation. Chimeric genes can be generated to produce therapeutic
 XX CC proteins with enhanced activity and vaccines can be produced. The
 XX CC present sequence is subtilisin E DNA, used to generate single-stranded
 XX CC DNA template.
 XX Sequence 1180 BP; 339 A; 262 C; 269 G; 276 T; 34 other;

Alignment Scores:

Pred. No.:	1.2e-14	Length:	1180
Score:	337.50	Matches:	117
Percent Similarity:	41.97%	Conservative:	66
Best Local Similarity:	26.83%	Mismatches:	156
Query Match:	6.01%	Indels:	97
DB:	22	Gaps:	17

US-09-830-837-6 (1-1052) x AAD19518 (1-1180)

QY	48	SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys	67
Db	113	AGCAGTACA-----GAAAGAATAATATGTCGGATT	148
QY	68	AlaArgAsnSerPheIleSerSerAlaLeuLysSerGluValAspAsnTyrArgIle	87
Db	149	CACAAATGATGCCATGAGTTCGCCCAAGAAAGGATGTTATT	193
QY	88	IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys	107
Db	193	-----	193
QY	108	GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal	126
Db	194	---TCTCAAAAGCGCGAAGGTTCAAAAGCAATTTAAGTATGTTAAACGCGCGCAGCA	250
QY	127	ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro	146
Db	251	ACATTGGATGAAAAGCTGTAAAGAAATGAAA-----AAAGATCCGACGGTGGCA	301
QY	147	CysAsnGluThrArgTrpSerGlnLysTyrSerArgProLeuArgAlaSer	166
Db	301	-----	301
QY	167	LeuSerLeuGlySerGlyPheThrHisAlaThrGlyArgHisSerSerArgArgLeuLeu	186
Db	302	-----TATGTGAAGAAGATCATATATGCATATGCAATATGCG	337
QY	187	ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuThrGlnMetGly	206
Db	338	CAATCTGTCTTATGGCATTTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAGGC	394
QY	207	TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro	226
Db	395	TACACAGGCTCTAACGTAAAGTAGCTGTATGACAGCGGAATGACTTCTCTCATCCT	454
QY	227	HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp	244
Db	455	GACTTA---AACGTGAGCGCGGAGCAAGCTTCGTACCTTCTGAAACAAACCCATACCAG	511
QY	245	AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----	262
Db	512	GACGGCAGTCTCAGCGTACGCATGTAGCGGTACGATTCGCGCTCTTAATACTCAATC	571
QY	263	---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn	281
Db	572	GGTGTCTGGCGGTAGCCCAAGCGCATCATATTATGTCAGTAAAGAGTCTTGATTCAACA	631
QY	282	GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle	301
Db	632	GGAAGCGGCAATATATGATTAATAAGCGCATTCAGTGGGCCATTTCCAAACAATATG	691
QY	302	AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal	319
Db	692	GATGTTATCAACATGAGCCTTGGCGGACCTTACTGGTCTACAGCGCTGAAACACGTCGT	751
QY	320	AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp	339
Db	752	GACAAAGCC-----GTTCCAGCGGTATCGTGTGCTGCCCGCAGCGGAAACGAA	802
QY	340	GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly	357
Db	803	GGTTTCATCGGAAGCACAGCAGTGGCTACCTTGCAAAATATCTCTTACTATTGCA	862


```
Qy 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
   |||
   |||:
   |||:
Db 923 GGTTCATCCGGAACACACAGTCGGCTACCTGCAGAAATATCTTCTACTATTGCA 982
   |||
   |||:
   |||:
Qy 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThr 377
   |||
   |||:
   |||:
Db 983 GTAGTCCGGTAAACACAGCAGACCAAGAGCTTCATCTCCAGCGAGTTCT----- 1036
   |||
   |||:
   |||:
Qy 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
   |||
   |||:
   |||:
Db 1037 ---GAGCTT-----CATGTGATGGCTCTCGCGTGCTCC 1066
   |||
   |||:
   |||:
Qy 398 ValArgGlySerGlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
   |||
   |||:
   |||:
Db 1067 ATCCAAAGCACACTTCCTGGAGGCACCTACGGCGCTTATAACGGAACGTCATCGCGCACT 1126
   |||
   |||:
   |||:
Qy 418 ProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
   |||
   |||:
   |||:
Db 1127 CCTCACGTTCCGGAGCAGCAGCGTTAATTCTTCT-----AAGCACCGCACTGG 1177
   |||
   |||:
   |||:
Qy 438 AsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsn 457
   |||
   |||:
   |||:
Db 1178 ACAAACCGCAAGTCCGTGATCGTTTAGAAGCACTGCAACATATCTTGA-----AAC 1231
   |||
   |||:
   |||:
Qy 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAla-TyrGlnIleLeuAsnSe 477
   |||
   |||:
   |||:
Db 1232 TCTTCTACTATGAAAGAGGTTAATCAACGTACACAGCAGCTGCACAAAT-----AATAG 1285
   |||
   |||:
   |||:
Qy 477 rTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMe 497
   |||
   |||:
   |||:
Db 1286 TAAAGAGAGCAGGTT-----CCTCCATACCTGCTTCTTTTATTGTCAGCAT-- 1334
   |||
   |||:
   |||:
Qy 497 tTrrProTyrCysSerGlnProIleTyrTyrGlyGlyMetProThrValValAsnValTh 517
   |||
   |||:
   |||:
Db 1335 -----CCTGATGTTCCGGCGCATTC 1354
   |||
   |||:
   |||:
Qy 517 rIleLeuAsnGlyMet-----GlyValThrGlyArgI1 528
   |||
   |||:
   |||:
Db 1355 TCTTCTTCTCCGCATGTTGAATCCGTTCCATGATCGACGGATGGCTGCTCTGAAATC 1414
   |||
   |||:
   |||:
Qy 528 eValAspLysProAspTrpGlnProTyrLeuProGlnAsnGly 542
   |||
   |||:
   |||:
Db 1415 TTCACAAGACCGGAGGATCAACCTGCTCAGCCCCCGTCACGGC 1457
   |||
   |||:
   |||:
```

Search completed: June 2, 2003, 02:58:04
Job time : 775 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 02:58:17 ; Search time 973 Seconds
(without alignments)
1457.842 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNLTWLLVLLCGKKH.....PRVKRPLMQVHPKTPSV 1052

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 845702 seqs, 674182571 residues

Total number of hits satisfying chosen parameters: 1691404

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-O/cgn2_1/USPTO.spool/US09830837/runat_23052003_181926_8115/app.query.fasta_1.1223
-DB=published.Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=25 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09830837 -CGN_1_1_201 @runat_23052003_181926_8115
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT
-DSPBLOCK=100 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10
-XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5617	100.0	4338	10	US-09-891-711-3 Sequence 3, Appli
2	5487	97.7	4198	10	US-09-891-711-5 Sequence 5, Appli
3	734	13.1	467	9	US-09-918-995-12598 Sequence 12598, A
4	393	7.0	300	10	US-09-960-352-8775 Sequence 8775, Ap

5	337.5	6.0	1074	10	US-09-837-235-14 Sequence 14, Appl
6	335.5	6.0	1140	10	US-09-920-118-13 Sequence 13, Appl
7	332.5	5.9	1146	10	US-09-920-118-15 Sequence 15, Appl
8	331	5.9	1497	7	US-08-322-678-6 Sequence 6, Appli
9	331	5.9	1497	7	US-10-033-325-1 Sequence 1, Appli
10	331	5.9	1497	10	US-09-060-854B-1 Sequence 1, Appli
11	320	5.7	1977	12	US-10-090-624-11 Sequence 11, Appl
12	302	5.4	1236	12	US-10-090-624-2 Sequence 2, Appli
13	302	5.4	1962	12	US-10-090-624-15 Sequence 15, Appl
14	285	5.1	4765	12	US-10-090-624-5 Sequence 5, Appli
15	282	5.0	1140	7	US-08-322-678-11 Sequence 11, Appl
16	279	5.0	1140	7	US-08-322-678-12 Sequence 12, Appl
17	261.5	4.7	576	10	US-09-920-118-21 Sequence 21, Appl
18	260.5	4.6	3788	9	US-09-927-827-33 Sequence 33, Appl
19	248.5	4.4	1485	10	US-09-974-300-1938 Sequence 349, Ap
20	247	4.4	497	10	US-09-783-590-349 Sequence 349, Ap
21	241.5	4.3	1329	10	US-09-974-300-1934 Sequence 1934, Ap
22	229.5	4.1	1330	10	US-09-966-921A-5 Sequence 5, Appli
23	228	4.1	1306	10	US-09-966-921A-1 Sequence 1, Appli
24	227.5	4.1	840	9	US-10-209-812-1 Sequence 1, Appli
25	225.5	4.0	1397	10	US-09-974-300-1933 Sequence 1933, Ap
26	225	4.0	1971	10	US-09-974-300-1935 Sequence 1935, Ap
27	221	3.9	522	9	US-09-824-893A-71 Sequence 71, Appl
28	218	3.9	429	9	US-09-918-995-35748 Sequence 3748, A
29	217	3.9	522	9	US-09-824-893A-94 Sequence 94, Appl
30	217	3.9	3650	9	US-09-927-827-31 Sequence 31, Appl
31	216	3.8	522	9	US-09-824-893A-55 Sequence 55, Appl
32	215	3.8	522	9	US-09-824-893A-112 Sequence 112, App
33	212.5	3.8	3825	10	US-09-974-300-1925 Sequence 1925, Ap
34	212	3.8	522	9	US-09-824-893A-107 Sequence 107, App
35	211	3.8	522	9	US-09-824-893A-97 Sequence 97, Appl
36	211	3.8	522	9	US-09-824-893A-128 Sequence 128, App
37	210	3.7	522	9	US-09-824-893A-57 Sequence 57, Appl
38	208	3.7	522	9	US-09-824-893A-45 Sequence 45, Appl
39	208	3.7	522	9	US-09-824-893A-102 Sequence 102, App
40	207	3.7	522	9	US-09-824-893A-84 Sequence 84, Appl
41	206	3.7	522	9	US-09-824-893A-74 Sequence 74, Appl
42	206	3.7	522	9	US-09-824-893A-104 Sequence 104, App
43	206	3.7	522	9	US-09-824-893A-110 Sequence 110, App
44	205	3.6	522	9	US-09-824-893A-123 Sequence 123, App
45	204	3.6	522	9	US-09-824-893A-38 Sequence 38, Appl

ALIGNMENTS

RESULT 1
US-09-891-711-3
; Sequence 3, Application US/09891711
; Patent No. US20020082404A1
; GENERAL INFORMATION:
; APPLICANT: Pharmacia AB
; TITLE OF INVENTION: Promoter Sequences
; FILE REFERENCE: 00130
; CURRENT APPLICATION NUMBER: US/09/891,711
; CURRENT FILING DATE: 2001-06-26 *to date*
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 4338
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (497)..(3655)
US-09-891-711-3

Alignment Scores:			
Pred. No.:	0	Length:	4338
Score:	5617.00	Matches:	1052
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0

Db	1047	CTCAGTGAAGCATCCACACTTCAAGAAATGTGAAGGAGAGAAACCAACTGGACCAATGAG	1106	QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProAspSer	620
QY	241	ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet	260	Db	2187	CATACTTCCACAGTGAAGCTTCCCATTAAGGTGAAGATCATTTCCACCCCTCTCGGAGC	2246
Db	1107	CGGACCTCGATGATGGCTGGCCATGGCACATTTGTCGAGGTGTGATTGCCAGCATG	1166	QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuAtrGlyTrpProGlyTyrPheProAsp	640
QY	261	ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn	280	Db	2247	AAGAGAGTCTCTGGGAGCAGTATCAACAACCTCCGCTACCCGCCAGCTACTTCCACAG	2306
Db	1167	AGGAGTGCACAGGATTTGCCCGAGATGCAGAGCTGCACATCTTCCGGGTCTTTACCAAC	1226	QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
QY	281	AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys	300	Db	2307	GACAACTTGGCATGAAGATGATCTTTAGACTGGAATGGCGCATGTGCCACACCAAT	2366
Db	1227	AATCAGGTGCTTACACATCTTGTTTTGGACGCTTTCACATATGCCATCTTAAGAGAG	1286	QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
QY	301	IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp	320	Db	2367	TTGAGGACATGTACACGACCTGCGCAGCATGGCTACTTCTGAGAGGTGCTCGGTGCC	2426
Db	1287	ATTGATGTTAAACCTTAGCATCGCGGCCCTGACTTCATGGATCATCCCTTTGTGAC	1346	QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
QY	321	LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly	340	Db	2427	CCATTCACGTGCTTGATGCTACACATGATGGCACCTTCTCATGTTGATGATGAAGAA	2486
Db	1347	AAGGTGGGAATTAACAGCTAACAAATGTAATCATGGTTCTGCTATCGGCAATGATGA	1406	QY	701	GlnTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720
QY	341	ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly	360	Db	2487	GAGTACTTCCAGAGGAGATTGCCAAGCTGAGGAGGAGCTGGCAATGGCTTTCCTC	2546
Db	1407	CCTCTTTATGGCACTCTGAATAACAGCTGATCAGATGGATGATGGAGTGGGTGGC	1466	QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
QY	361	IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrTrpGluLeu	380	Db	2547	GTCACTTCTAGTCTGCTGACACACTTCTGTTATCAGAAAGTGAAGTATTTACGATGA	2606
Db	1467	ATTGACTTTGAAGATAACATCGCCCGCTTTCTTCCAGGGAAATGACTACCTGGGAATA	1526	QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu	760
QY	381	ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly	400	Db	2607	AACACAGGAGGTGGGTGGATGCCAGTACTGGAGGAGGACCAACATCCAGCTCTGAACGAG	2666
Db	1527	CCAGGAGGCTATGTCGCTGGAACCTGACATGTCACCATGTTGGTGGCGAGTGGCGGT	1586	QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
QY	401	SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal	420	Db	2667	CTGCTGCTGCTGGGAACATGGGTTCAGCGATGGCTTTATGAAGGGAGTGTCCCTG	2726
Db	1587	TCGGTGTGAAGGGGCTGCGGGCACTCTCAGGACCACTGCTGCTTCCCGAGTGGTT	1646	QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
QY	421	AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla	440	Db	2727	GGCAATCATGACATGATTTATGATCGGATGAGCATGCCAAGTCTTCCAGAAGATGGT	2786
Db	1647	GCTGGGCTGTACCTTGTAGTAGCACAGTGCAGAGGGAGCTAGTGAATCTGCC	1706	QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
QY	441	SerMetLysGlnAlaLeuIleAlaSerAlaArgAlaLeuProGlyValAsnMetPheGlu	460	Db	2787	GTGTGATCAGACAGACTTTCAAGGACCAAGGATTTGAGGTCTTAAACAAGAGACAGCA	2846
Db	1707	AGTGTGAAGCAGCCCTGATTCATCAGCCGCGAGCTTCTGTTGTTACATGTTGAG	1766	QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
QY	461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480	Db	2847	GTGTTGAAATGTTCCCATTTTGGGCTTTATCAGATTCAGCTGAAGTGGGGCGCG	2906
Db	1767	CAAGCCATGGCAAGCTGATCTGCTGGAGCCTATCAGATCTCAGCAGCTACAACCA	1826	QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500	Db	2907	ATCGTGTGTATGGAGATTCCAATTGCTTGGATCAGATCAGACAGACAGAGGATTTGCTTT	2966
Db	1827	CAGCGAGCTTGATCTGATCAGCTACATCAGCTGAGTGTGCTACATGTGGCCTTAC	1886	QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
QY	501	CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn	520	Db	2967	TGGCTTCTGGATGCATCTTCTCAGTACATCATATGCGCTGAACCTCCAGCTCCAGCTCAGC	3026
Db	1887	TGTTCTCAGCCCATCTACTATGAGGAATGCCAACAATGTTTAATGTACACATCTTCAT	1946	QY	881	HisSerGlyAsnArgGlnArgProProSerGlyIleGlySerValThrProGluArgMet	900
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540	Db	3027	CATTTCAGGAAACCGGAGCGCCACCCAGTGGAGTGGTGGTGGCTTCTTGAAGAGTGT	3086
Db	1947	GGCATGGAGTACACAGGAAGATTTGATGAAGCTGAGTGGCGGCCCTATTATCCACAG	2006	QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTyrProTyrSerGly	560	Db	3087	GAAGGAAACCACTTCTCATCTACTCCAAAGGTCTTGGAGCCCATCTGGAGACCCAAAA	3146
Db	2007	AATGAGACACATGAGTGGCTTCTCTACTCTCAGTGTATGGCTTGTGGCTCAGGC	2066	QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
QY	561	TyrLeuAlaIleSerIleSerValThrLysAlaAlaSerTrpGluGlyIleAlaGln	580	Db	3147	CCTGGGCTCTTCCAGCCTGTCCACACTTGTCTATGGGCGCAAGCAGCAGCTTTGATGAG	3206
Db	2067	TACCTGGCCATCTCCATTTCTGTGACCAAGAGAGCTTCTCTGGAGAGGATTTGACAG	2126	QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600	Db	3207	ACTGGCCCACTGATCATCTTTGGAAACATCAGAAAGTGTCTTCCATTCACCTGGACAAAGTA	3266
Db	2127	GGTCACATCATGATCAGGTGGCTTCCCGCAGCAGAGCGAAGCAAAAAATTTGGTGGCGAG	2186				

QY 961 ValLeuProAspPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3267 GTGTTACCAACTTTTCATCGAATCGCCCTCAAGTCAGACCTTTGTCCCTCGAGAAAGT 3326
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnValGlyGln 1000
Db 3327 GGTGCTCGGACATTCCTGGAGGATCATGCTGCGCGCTACAAACAGAGGTGGCCAG 3386
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3387 ACCATCCCTGCTTTCCTCTCCGCGCATGGTGGCCCTCTTTGTGGTACAG 3446
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3447 ATCAGCAAGGCCAAAGCGCGGCGAAGCGAGGAGGCCCAAGGCGTCCACAGCTT 3506
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3507 ACACAGCAGACCCACCCACCAAGGACCCGCTCAGTG 3542

RESULT 3
US-09-918-995-12598
; Sequence 12598, Application US/09518995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12598
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(467)
; OTHER INFORMATION: n = A,T,C or G

US-09-918-995-12598

Alignment Scores:
Pred. No.: 1,93e-69 Length: 467
Score: 734.00 Matches: 138
Percent Similarity: 99.28% Conservative: 0
Best Local Similarity: 99.28% Mismatches: 1
Query Match: 13.07% Indels: 0
DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-918-995-12598 (1-467)

QY 710 LeuArgArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThr 729
Db 48 CTCGGAGGAGCGTGTGACACGCGCTCTCGCTCATCTTCAGTGACGTGGTATTACACT 107
QY 730 SerValMetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpTrpMetProAsp 749
Db 108 TCTGTTATGAGAAAGTGAAGTTTATGATGAAACACAAAGGCGAGTGTGATGCCGGAT 167
QY 750 ThrGlyGlyValAsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPhe 769
Db 168 ACCGGAGGAGCTAACATCCAGCTCTGAATGAGCTGCTGTGTGGAAACATGGGGTTC 227
QY 770 SerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSer 789
Db 228 AGCGATGGCTGTATGAGGGAGTTCACCTGGCCCAACCATGATGATATGCGTCA 287
QY 790 GlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThrPheLysAsp 809
Db 940 GTGTTATGAGAAAGTGAAGTTTATGATGAAACACAAAGGCGAGTGTGATGCCGGAT 167

Db 288 GGGTCGACGATCCGCGAAGTTTCCAGAAGATGGCGTCTGTATAACACAGACTTTTCAAGGAC 347
QY 810 GlnGlyLeuGluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGly 829
Db 348 CAGGATTGGAGTTTAAAGCAGGAAACAGCAGTTGTTGAAACGTCCTCCCATTTGGGA 407
QY 830 LeuTyrGlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsn 848
Db 408 CTTTATCAGATTCACGCTGAGGCGCGGATTGTACTGTATGGGACTCCAAT 464

RESULT 4
US-09-960-352-8775
; Sequence 8775, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 8775
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 38-LIB188-011-01-E1-B6

US-09-960-352-8775

Alignment Scores:
Pred. No.: 3,7e-33 Length: 300
Score: 393.00 Matches: 70
Percent Similarity: 94.81% Conservative: 3
Best Local Similarity: 90.91% Mismatches: 4
Query Match: 7.00% Indels: 0
DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x US-09-960-352-8775 (1-300)

QY 832 GlnIleProAlaGluGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAsp 851
Db 3 CAGCTCACAGCTGAGGCGCGCGCTGCTGCTGTATGGAGACTCCAACTGCTTGAC 62
QY 852 AspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuGlnTyrThrSer 871
Db 63 GACAGTCACCGCGCAGAGGACTGCTTCTGGCTCCTGGATGCACTCCTGCACTTCACGTCG 122
QY 872 TyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGly 891
Db 123 TATGGGTGACGCGCGCGCGCGCTCAGCCACTCGGGGGCTCGGAGGCGCGCGCGCGGGA 182
QY 892 AlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyr 908
Db 183 GCAGGCTCGCGACGCCCGGAGGAGGTGGAAGGGAACACCTGCATCGGTAC 233

RESULT 5
US-09-837-235-14
; Sequence 14, Application US/09837235
; Patent No. US20020061549A1
; GENERAL INFORMATION:
; APPLICANT: Marshall, Christopher
; APPLICANT: Hoffman, Alexander
; APPLICANT: Errico, Joseph
; APPLICANT: Marshall, Paul
; TITLE OF INVENTION: STABILIZED PROTEINS
; FILE REFERENCE: 9725-005-999
; CURRENT APPLICATION NUMBER: US/09/837,235
; CURRENT FILING DATE: 2001-04-18
; PRIOR APPLICATION NUMBER: PCT/US00/28595
; PRIOR FILING DATE: 2000-10-16

```
; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-837-235-14

Alignment Scores:
Pred. No.: 2,52e-26 Length: 1074
Score: 337.50 Matches: 117
Percent Similarity: 41.97% Conservative: 66
Best Local Similarity: 26.83% Mismatches: 156
Query Match: 6.01% Indels: 97
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-837-235-14 (1-1074)
QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
DB 25 ACAGTACA-----GAAAGAAATACATTGTCGGATT-----AAA 60
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAlaPheAsnTrpArgIle 87
DB 61 CAGACAATGAGTCCATGAGTTCCGCCAAGAAAGAGATGTTATT----- 105
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
DB 105 ----- 105
QY 108 GluLysGlnLysAlaGlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgVal 126
DB 106 ---TCTGAAAGAGCGCAAGGTTCAAAGCAATTTAAGTATGTAACGGCGCGCAGCA 162
QY 127 ThrProGlnArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValPro 146
DB 163 ACATTGGATGAAAGCTGTAAGAAATTCGAAA-----AAGATCCGACGGCTTGA 213
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
DB 213 ----- 213
QY 167 LeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
DB 214 -----TATGTGAAGAGATCATATGTCACATGAATACGG 249
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
DB 250 CAATCTGTCTTATGCGATTCTCAA---ATTAAAGCGCGGCTCTTCACTCTCAAGCC 306
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
DB 307 TACACAGGCTCTAACGTTAAAGTAGCTGTTATCAGACGGGAATGACTCTTCTCATCT 366
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
DB 367 GACTTA---AACGTACAGAGCGGAGCAAGCTTGCTACCTCTGAAACAAACCCATACCAG 423
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
DB 424 GAGCGAGTTCTCAGCGTACGATGTAGCCGGTACGATTGCGGCTCTTAATACTCAATC 483
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
DB 484 GGTGTTCTGGCGTTAGCCCAAGCCATCATATATAGTAAGTAAGTGTGATTGATTCACA 543
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIle 301
DB 544 GNAAGCGCCAAATATAGCTGGATTATATACCGCATTCAGTGGCCATTTCCACAAATAG 603
QY 302 AspValLeuAsnLeuSerIleGlyClyProAspPheMetAsp-----HisProPheVal 319

; PRIOR APPLICATION NUMBER: 60/159,763
; PRIOR FILING DATE: 1999-10-15
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 14
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-837-235-14

Alignment Scores:
Pred. No.: 4.53e-26 Length: 1140
Score: 335.50 Matches: 126
Percent Similarity: 39.18% Conservative: 66
Best Local Similarity: 25.71% Mismatches: 169
Query Match: 5.97% Indels: 129
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-920-118-13 (1-1140)
QY 1 MetLysLeuValAsnIleTrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
DB 4 ATGAGGAAAGAGTTTGGCTGGAGTGTGACGCGCTTAATG----- 48
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40

RESULT 6
US-09-920-118-13
; Sequence 13, Application US/09920118
; Patent No. US20020102734A1
; GENERAL INFORMATION:
; APPLICANT: Menzel, Rolf
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 13
; LENGTH: 1140
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1140)
US-09-920-118-13

Alignment Scores:
Pred. No.: 4.53e-26 Length: 1140
Score: 335.50 Matches: 126
Percent Similarity: 39.18% Conservative: 66
Best Local Similarity: 25.71% Mismatches: 169
Query Match: 5.97% Indels: 129
DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-920-118-13 (1-1140)
QY 1 MetLysLeuValAsnIleTrpLeuLeuLeuValValLeuLeuCysGlyLysLysHis 20
DB 4 ATGAGGAAAGAGTTTGGCTGGAGTGTGACGCGCTTAATG----- 48
QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
```


235 ACATTGATGAAAGAGCTGTAAAGAAATGAAA-----AAAGATCGGCGCTTCCA 285
QY 147 CysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSer 166
Db 285 ----- 285
QY 167 LeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeu 186
Db 286 -----TATGCGAAGAGATCATATTGCACATGAATATGCG 321
QY 187 ArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGly 206
Db 322 CAATCTGTCCTTATGCGATTTCTCA-----ATTAAAGCGCGGCTCTTCACTCTCAAGGC 378
QY 207 TyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisPro 226
Db 379 TACACAGCTCTTACCTAAAGTAGCTGTTCGACAGCGGAATGACITCTCTCATCT 438
QY 227 HisPheLysAsnValLysGluArgThrAsnTrpThrAsnGluArgThr-----LeuAsp 244
Db 439 GACTTA---AACGTGACAGGCGGAGCAAGCTTCTGACCTTCTGAAACAACACCATACCAG 495
QY 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 496 GACGCGAGTTCTCACGGTACGATGTAGCGGTACGATGCGCTCTTAATACTCAATC 555
QY 263 ---CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsn 281
Db 556 GGTGTTCTGCGCTAGCGCAAGCGCATATTATATGCAAGTAAAGTGTGTGATTCACAA 615
QY 282 GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysIle 301
Db 616 GGAAGCGGCCAATATAGCTGATTTAAGCGCATTTAGTGGCCATTTCCAAACAATAG 675
QY 302 AspValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheVal 319
Db 676 GATGTTATACATAGAGCTTGGCGGACCTTACTGCTTCTACAGCGCTGAAACAGTCGT 735
QY 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 736 GACAAAGCC-----GTTTCCAGCGGTATGCTGCTGCTGCGGCGGCAACGAA 786
QY 340 GlyPro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGly 357
Db 787 GGTTCATCCGAGCAGCAGCAGCTGCGGTACCGCTGACCAAAATATCTCTTACTATGCA 846
QY 358 ValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 377
Db 847 GTAGTGCGGTAAACAGCAGCAACCAAGAGCTTCAATCTCCAGCGCAGGTCT----- 900
QY 378 TrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGly 397
Db 901 ---GAGCTT-----GATGTGATGGCTCTCGCGGTGCC 930
QY 398 ValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSer 417
Db 931 ATCCAAAGCACACTTCTGAGGCGACTTACGGCGCTTATAACGGAAGCTCATGCGGACT 990
QY 418 ProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuVal 437
Db 991 CCTCAGTTGCCGAGCAGCAGCTTATTTCTTCT-----AAGCAGCGACTTGG 1041
QY 438 AsnProAlaSerMetLysGlnAlaLeuLeuAlaSerAlaArgArgLeuProGlyValAsn 457
Db 1042 ACAACGCGCAGTCCGTGATCGTTTGAAGACACTGCACATATCTTGA-----AAC 1095
QY 458 MetPheGluGlnGlyHisGlyLysLeuAspLeuArgAlaTyrGln 473
Db 1096 TCTTCTACTATGGAAGAGGTATATCAACGTACAGCAGCTGCACAA 1143
RESULT 8
US-08-322-678-6
; Sequence 6, Application US/08322678

Publication No. US20030077807A1
GENERAL INFORMATION:
APPLICANT: Graycar, Thomas P
APPLICANT: Bott, Richard R
APPLICANT: Wilson, Lori J
TITLE OF INVENTION: Subtilisin Variants
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genencor International, Inc
STREET: 180 Kimball Way
CITY: So. San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/322,678
FILING DATE: 13-Oct-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Horn, Margaret A.
REGISTRATION NUMBER: 33,401
REFERENCE/DOCKET NUMBER: GC235-2
TELEPHONE: (415) 742-7536
TELEFAX: (415) 742-7217
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 1497 base pairs
TYPE: nucleic acid
STRAINEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-322-678-6

Alignment Scores:
Pred. No.: 2,12e-25 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
Gaps: 19
DB:

US-09-830-837-6 (1-1052) x US-08-322-678-6 (1-1497)
QY 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuLysAspArgLeuGluLys 27
Db 45 ATAATCTCTATTGGTTATCTCTCAATGAAAGAGAGAGAGGATAAGAGCTGAGAGC 104
QY 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGATGATGATGATTTGCTGTTTCTTTTACG-----TTA 143
QY 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTACGATGGCTTCGCGACACATCTCTCCCGAGCGCGGAGGAAATCAACGGG 203
QY 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATTGTCGGGTT-----AACAGACATGAGCAGC 245
QY 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsn 92
Db 246 ATGACGCGCGCTAGAGAGAGATGTCATT----- 275
QY 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287

```
Qy 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAGTGCAAAAGCAATTCAAATATGTAGCGCAGCTTCAGTCACATTAAACGAAAA 347
Qy 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAGAAAGTAAATGAAA-----AAAGACCCGAGCGTCCT----- 383
Qy 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySer 171
Db 383 ----- 383
Qy 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCAGTACGACATCGGTACGCGCAGTCGTCCTTAC 434
Qy 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsn 211
Db 435 GCGGTATCAAA---ATTAAGCCCTGCTCTGCACTCTCAAGCTACACTGATCAAT 491
Qy 212 ValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys----- 229
Db 492 GTTAAGTACGGTTATCGACACGGGTATCGATCTCTCATCTGATTTAAGGTAGCA 551
Qy 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 AGCGGAGCCAGCATGGTCTCTCTGAACAAAT-----CCTTTCCAGAC 596
Qy 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 ACAACTCTCAGGAACCTACGTTGCGGCGACAGTTGCGGCTCTTAATAACTCAATCGGT 656
Qy 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGCGGTGCGCAACGCGATCATCTTACGCTGTAAGACTTCTCGGTGTCAGCGT 716
Qy 283 ValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCGGCGCAATPACAGCTGGATCATTAACGAATCGAGTGGCGATCGCAACAATATGGAC 776
Qy 303 ValLeuAsnLeuSerIleGlyLysProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAACTATGACCTCGCGGACCTCTGGTTCGTCTGCTTTAAAGCGGCGGTGAT 836
Qy 321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAAGCC-----GTTCATCCGCGCTGCTAGTCTTCGCGCAGCCGCTACGAAGC 887
Qy 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCAGCTCAAGCAGTGGGCTACCTGGTAAATACCTCTCTGTCATTGCAGTA 947
Qy 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GCGCGTGTTCAGCAGCAGCAACCAAGAGCATCTTCTCAAGCTAGGACCT----- 998
Qy 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GACCTT-----GATGTCATGCGACCTGCGCTATCTATC 1031
Qy 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValala 416
Db 1032 CAAAGCAGCGTCTCTGGAACAATAACGGG-----CGGTACACGCTAGCTCAATGGCA 1085
Qy 417 SerProValValAlaGlyAlaValThrLeuValSer 429
Db 1086 TCTCCGACGCTTCGCGGCGGCTGCTTTGATCTTCT 1124
```

RESULT 9

US-10-033-325-1

; Sequence 1, Application US/10033325

; Publication No. US2003007322A1

; GENERAL INFORMATION:

```
; APPLICANT: Poulou, Ayrookaran J.
; APPLICANT: Schellenberger, Volker
; APPLICANT: Kellis, Jr., James T.
; APPLICANT: Paech, Christian
; APPLICANT: Nadherny, Joanne
; APPLICANT: Naki, Donald P.
; APPLICANT: Collier, Katherine D.
; APPLICANT: Caldwell, Robert M.
; APPLICANT: Baek, Andre C.
; TITLE OF INVENTION: Multiply-Substituted Protease Variants
; FILE REFERENCE: GC502-2-C1
; CURRENT APPLICATION NUMBER: US/10/033,325
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: US 09/178,155
; PRIOR FILING DATE: 1998-10-23
; PRIOR APPLICATION NUMBER: US 08/956,323
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,564
; PRIOR FILING DATE: 1997-10-23
; PRIOR APPLICATION NUMBER: US 08/956,324
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: Bacillus amyloliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
; US-10-033-325-1

Alignment Scores:
Pred. No.: 2,12e-25 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
DB: Gaps: 19

US-09-830-837-6 (1-1052) x US-10-033-325-1 (1-1497)
Qy 8 LeuLeuLeuValValLeuLeuCysGlyLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGTCTATGTTATCTCAATGAAAAAGAGAGAGATAAGAGTGAAGGC 104
Qy 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGTATGATCATGTTGCTTTGCTTTAGCG-----TTA 143
Qy 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTACGATGCGGTGCGCAGCATCTCTCCCGCGCGGGAATCAACGCG 203
Qy 53 GluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPhe 72
Db 204 GAAAGAAATATATGTCGGGTTT-----AAACAGACAATGAGCAGC 245
Qy 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
Db 246 ATGAGCGCGCTAAGAAGAAAGATGTCATT----- 275
Qy 93 ProSerSerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAla 112
Db 276 -----TCTGAAAAGGC 287
Qy 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArgLys 131
Db 288 GCGAAGTGCAAAAGCAATTCAAATATGTAGCGCAGCTTCAGTCACATTAAACGAAAA 347
Qy 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 348 GCTGTAAGAAAGTAAATGAAA-----AAAGACCCGAGCGTCCT----- 383
```

```
152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGlySer 171
383 -----
172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuAlaIleProArg 191
384 -----TACGTTGAAGAAGATCACGTAGCACATCGTACGCGCAGTCCGTGCCTTAC 434
192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrThrGlyAlaAsn 211
435 GCGGTATACAAA--ATTAAGCCCTGCTCTGCACTCTCAAGCTACACTGGATCAAT 491
212 ValArgValAlaValPheAspThrGlyLeuSerGlnLysHisProHisPheLys----- 229
492 GTTAAGTAGCGGTATTCACAGCGGTATCTCTCATCTCTGATTTAAAGGTAGCA 551
230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
552 ACGGAGCCAGCATGTTCTTCTGAACAAAT-----CCTTTCCAGAC 596
246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
597 ACAACTCTCTACGGAACCTCACGTTCCGCGCACAGTTGGGCTCTTAATAACTCAATCGGT 656
263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
657 GTATTAGCGGTTCGCCAAGCCATCATCTTACGCTGTGTAAGTTCTCGGTGTCACGGT 716
283 ValSerTrpThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLysIleAsp 302
717 TCGGCCCAATACAGCTGGATCATTAAGGAATCGAGTGGCGATCGCAACAATATGGAC 776
303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
777 GTATTAACTAGAGCTCGCGGACCTCTGCTGCTGCTGCTTTAAAGGGCGAGTTGAT 836
321 LysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGly 340
837 AAAGCC-----GTTGCATCCGGCTGCTAGTCTGTTGGCGGCGCCGTAACGAGGC 887
341 Pro-----LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
888 ACTTCCGGCAGCTCAGCACAGTGGCTACCCGCTGTAATACCTTCTGTCATTCAGTA 947
359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
948 GCGCTGTGACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
379 GluLeuProGlyGlyTrpGlyArgMetLysProAspIleValThrTrpGlyAlaGlyVal 398
999 GAGCTT-----GATGTCATGGCCACCTGGCGGTATCTATC 1031
399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
1032 CAAGCAGCGTTCTCTGGAACAAATACGGG-----GCTACACGGTAGTCAATGGCA 1085
417 SerProValAlaGlyAlaValThrLeuValSer 429
1086 TCTCCGCGAGTTCGCGAGCGGCTGCTTTCATCTTTCT 1124

RESULT 10
US-09-060-854B-1
; Sequence 1, Application US/09060854B
; Patent No. US20020081703A1
; GENERAL INFORMATION:
; APPLICANT: Estall, David Aaron
; TITLE OF INVENTION: Human Protease and Use of Such Protease for Pharmaceutical
; TITLE OF INVENTION: Applications and for Reducing the Allergenicity of No. US2002008
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: GC532
; CURRENT APPLICATION NUMBER: US/09/060.854B
; CURRENT FILING DATE: 1998-04-15
```

```
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1497
; TYPE: DNA
; ORGANISM: B. amyloliquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (96)...(1245)
US-09-060-854B-1

Alignment Scores:
Pred No.: 2, 12e-25 Length: 1497
Score: 331.00 Matches: 124
Percent Similarity: 40.40% Conservative: 59
Best Local Similarity: 27.37% Mismatches: 146
Query Match: 5.89% Indels: 124
Db: Gaps: 19

US-09-830-837-6 (1-1052) x US-09-060-854B-1 (1-1497)
Qy 8 LeuLeuLeuValValLeuLeuCysGlyLysLysHisLeuGlyAspArgLeuGluLys 27
Db 45 ATAATCTGCTCTATGTTGTTATCTGCAATGAAAAAGAGAGATAAGAGTGAAGAGC 104
Qy 28 LysSer-----PheGluLysAlaProCysProGlyCysSerHisLeu 41
Db 105 AAAAAAGTATGATCATGTTGCTGTTGCTTTAGCG-----TTA 143
Qy 42 ThrLeuLysValGluPheSerSerThrValVal----- 52
Db 144 ATCTTTAGCATGGGTCGCGACACATCTCTCCAGCGCGAGGAAATCAACAGGG 203
Qy 53 GluTrpGluTrpIleValAlaPheAsnGlyTrpPheThrAlaLysAlaAsnSerPhe 72
Db 204 GAAAGAAATATATTTTCGGGTT-----AAACAGCAATGAGCAGC 245
Qy 73 IleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsn 92
Db 246 ATGACGCCCTAAGAAAGATGTCATT----- 275
Qy 93 ProSerSerAspTrpProSerAspPheGluValIleGlnIleLysGluLysAla 112
Db 276 -----TCTGAAAAAGGC 287
Qy 113 GlyLeuLeuThrLeuGlu---AspHisProAsnIleLysArgValThrProGlnArg 131
Db 288 GGAAGATGCAAAAGCAATCAATATATGACCGCAGCTTCAGTCACTTAACGAAAAA 347
Qy 132 ValPheArgSerLeuLysTrpAlaGluSerAspProThrValProCysAsnGluThr 151
Db 348 GCTGTAAAGAAATTTGAAA-----AAAGACCCGAGCGTCGCT----- 383
Qy 152 TrpSerGlnLysTrpGlnSerSerArgProLeuArgArgAlaSerLeuSerLeuGly 171
Db 383 ----- 383
Qy 172 GlyPheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuAlaIleProArg 191
Db 384 -----TACGTTGAAGAAGATCACGTAGCACATCGTACGCGCAGTCCGTGCCTTAC 434
Qy 192 GlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrThrGlyAla 211
Db 435 GCGGTATACAAA--ATTAAGCCCTGCTCTGCACTCTCAAGCTACACTGGATCAAT 491
Qy 212 ValArgValAlaValPheAspThrGlyLeuSerGlnLysHisProHisPheLys----- 229
Db 492 GTTAAGTAGCGGTATTCACAGCGGTATCTCTCATCTCTGATTTAAAGGTAGCA 551
Qy 230 -----AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAspAsp 245
Db 552 ACGGAGCCAGCATGTTCTTCTGAACAAAT-----CCTTTCCAGAC 596
```

```
Qy 246 GlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu----- 262
Db 597 AACAACTCTCAGGAACACTACCTGCGGCACACGCTTAACTAACTCAATCGGT 656
Qy 263 CysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGln 282
Db 657 GTATTAGGCTTCCGCGCAGCGCATACCTTACGCTGTAAAGCTTCCGGTGTGACGGT 716
Qy 283 ValSerTyrThrSerTyrPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp 302
Db 717 TCCGGCCAATACAGCTGGATCATTAACGAATCGATGGCGGATCCAAACATATGGAC 776
Qy 303 ValLeuAsnLeuSerIleGlyProAspPheMetAsp-----HisProPheValAsp 320
Db 777 GTATTAAATAGATGACCTCGCGGACCTTCTGCTCTCTGCTTTAAAGCGGCGAGTTGAT 836
Qy 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
Db 837 AAGACC-----GTTGATCCCGGCTCGGTAGTCTGCGCGACGCCGGAACGAAGGC 887
Qy 341 Pro-----LeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyVal 358
Db 888 ACTTCGGCGGAGCTCAAGCAGTGGGCTACCTCTGTAATACCTTCTGTCATTGCGATG 947
Qy 359 GlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrp 378
Db 948 GCGCTGTGTACAGCAGCAACCAAGAGCATCTTCTCAAGCGTAGGACCT----- 998
Qy 379 GluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal 398
Db 999 GAGCTT-----GATGTCATGGCACTGGCGGTATCATC 1031
Qy 399 ArgGlySer-----GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1032 CAAGCAGCGTCTCTGGAAACATACGGG-----GCGTACACGCTAGCTCAATGGCA 1085
Qy 417 SerProValValAlaGlyAlaValThrLeuLeuValSer 429
Db 1086 TCTCCGACGCTTCCGCGAGCGGTGCTTTGATTCTTTCT 1124
```

RESULT 11

```
US-10-090-624-11
; Sequence 11, Application US/10090624
; Patent No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090, 624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11
```

```
Alignment Scores:
Pred. No.: 4,99e-24 Length: 1977
Score: 320.00 Matches: 127
Percent Similarity: 40.72% Conservative: 64
Best Local Similarity: 27.08% Mismatches: 185
```

```
Query Match: 5.70% Indels: 93
DB: 12 Gaps: 22
US-09-830-837-6 (1-1052) x US-10-090-624-11 (1-1977)
Qy 199 AlaAspValLeuTrp---GlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPhe 217
Db 439 GCGATACCTCTGGAACCTCCCTCGGTACGAGGAAGCGTGTGGTGGCATGTC 498
Qy 218 AspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrp 237
Db 499 GATACGGGTATAGACGGGAACCCCGGATCTGAAGGGC-----AAGGTCATAGGCTGG 552
Qy 238 -----ThrAsnGluArgThrLeuAspGlyLeuHisGlyThrPheVal 253
Db 553 TAGCAGCGCTCAACGGCAGTCGACCCCTACGATGACCGAGGACACGACCCACGTT 612
Qy 254 AlaGlyValIleAlaSerMetArgGluCysGln-----GlyPheAlaProAsp 269
Db 613 GCGGGTATCTTCCGGAACCGGCGGTTAACTCCAGTACATAGCGCTCGCCCGGC 672
Qy 270 AlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPhe 289
Db 673 GCGAAGCTCTCGCGTCAAGGTCTCGGTGCGCAGCGTTCGGGAAGCGTCTCCACCATC 732
Qy 290 LeuAspAlaPheAsnTyrAlaIleLeuLysLys-----IleAspValLeuAsn 305
Db 733 ATCGCGGTCTTGACTGGGTCTGCGACCAAGCAAGTACGGGATAAGGTGTCATCAAC 792
Qy 306 LeuSerIleGly-----GlyProAspPheMetAspHisProPheVal 319
Db 793 CTCTCCTCGCTCTCTCCAGAGCTCGAGCAACCGACTCCCTCAGTCAGCGC---GTC 849
Qy 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 850 AACACGCTCGGAGCGC-----GGTATAGTACTGCTGCGTCCGCGCGGCAACAGC 900
Qy 340 GlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 901 GGGCGGAACACCTACACCGTCCGCTCACCCTCGCGCGAGCAAGGTTCATAACCGTCGT 960
Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 961 GCAGTTGACAGCAACGACACATCGCCAGCTTCTCCAGCAGGGA----- 1005
Qy 380 LeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyVal--- 398
Db 1006 ---CCGACCGCGGAGGAGGCTCAAGCCGGAAGTCTGTCGCCCGCGGTGACATCATA 1062
Qy 399 -----ArgGlySerGlyValLysGlyGlyCys-----ArgAlaLeu 410
Db 1063 GCGCGCGCGCGAGCGGAACACGATGGGCAACCGGATAAAGCTACTACCAAGGCC 1122
Qy 411 SerGlyThrSerValAlaSerProValAlaGlyAlaValThrLeuLeuValSerThr 430
Db 1123 TCTGGAACCAAGCATGGCCACCGCGAGCTTTCGGCGGTTCGCGCTCATC----- 1173
Qy 431 ValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAla 450
Db 1174 CTCAGCGCCACCGGAGCTGGACCCCGGACAAGGTGAAGACCGCCCTTCATCGAGACGCC 1233
Qy 451 ArgArgLeuProGlyValAsnMetPheGlu-----GlnGlyHisGlyLysLeuAspLeu 468
Db 1234 GACATAGTCGCCCCCAAGGAGATAGCGGACATCGCTACGCTCGGGTAGGGTAGAGTC 1293
Qy 469 LeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAlaSerLeu-----SerPro 486
Db 1294 TACAAGGCCATCAAG-----TACGAGCACTACGCCAAGCTCACTTCACCGGC 1341
Qy 487 SerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSerGlnProIleTyr 506
Db 1342 TCGTCCGCGGACAAGGGAAGCGCCACCCACACCTTCGACGCTAGCGGC----- 1389
```

```
QY 507 TyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThrGly 526
Db 1390 -----GCCACCTTCGTGACCGCCACCTCTACTGGGACACGGC----- 1428
QY 527 ArgIleValAspLysProAspTrpGlnProTyrLeu---ProGlnAsnGlyAspAsnIle 545
Db 1429 -----TCGAGCGACATCGACCTCTACCTCTACGACCCCAACGGAACGAGGT 1476
QY 546 GluValAlaPheSer-----TyrSerSerValLeu 555
Db 1477 GACTACTCTACACCGCTACTACGGTTCGAGAGGTCCGCTACTACAACCGACCGCC 1536
QY 556 TrpProTrpSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrp 575
Db 1537 GGAACCTGG-----ACGGTCAGGTCGTCACTAC 1566
QY 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595
Db 1567 AGGCGCGCGG-----AACTACCGGTGACGTGTCGACCGGAGGCGCTCAGCCAG 1620
QY 596 LysAsnGly-AlaGluGlnThrSerThrValLysLeuProIleLysValLysIleIlePr 615
Db 1621 TCGGCGCGCGCAACCGGAATCCAAC-----CCAACCCGAACCAACCGCGGCC 1671
QY 615 oThrProArgSerLysArgVal 623
Db 1672 ACCGACCGCGACCTTCACCGGTT 1696

RESULT 12
US-10-090-624-2
; Sequence 2, Application US/10090624
; Patent No. US2002013235A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Kiyoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 2, 02e-22 Length: 1236
Score: 302.00 Matches: 119
Percent Similarity: 43.78% Conservative: 64
Best Local Similarity: 28.47% Mismatches: 150
Query Match: 5.38% Indels: 87
DB: 12 Gaps: 19

US-09-830-837-6 (1-1052) x US-10-090-624-2 (1-1236)
QY 186 LeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMet 205
Db 7 TTAGAGGACTGGATGAGTCTGACGCTCAAGTTATG---GCAACTACGTTTGGAACTTG 63
QY 206 GlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHis 225
Db 64 GGATATGATGGTCTCTGGGAATCACAATAGGAATATATGACACGGAATGACGCTCTCTCAT 123
```

```
QY 226 ProHisPhelys---AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAsp 244
Db 124 CCAGATTCACAGCAAAAGATTAATGGGTGGGTAGTATTCTCAATGGTAGAGTATACCA 183
QY 245 ---AspGlyLeuGlyHisGlyThrPheValAlaGlyVal-----IleAla 258
Db 184 TAGCATGACCATGACATGAACTCATGTAGTCTCAATAGCAGCTGGTACTGGAGCAGCA 243
QY 259 SerMetArgLysGlnGlnGlnPheAlaProAspAlaGluLeuHisIlePheArgValPhe 278
Db 244 AGTAATGGCAAGTACAGGAAGTGGCTCCAGGAGCTAAGCTGGCGGGAATTAAGAGTTCTA 303
QY 279 ThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeu 298
Db 304 GGTGCGCATGTTCTTCAAGACATATCTATAATTAAGGAGTGTGAGTGGCGGTGAT 363
QY 299 LysLys-----IleAspValLeuAsnLeuSerIleGly----- 309
Db 364 AACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTTGTGTTCAAGCCAGAGCTCA 423
QY 310 ---GlyProAspPheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsn 328
Db 424 GATGTACTGACGCTCTAAGTCAGCT---GTTAATGACGCTGGGATGCT--- 471
QY 329 AsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsn 348
Db 472 GGATTAGTGTGTGGTTCGCTCGCAACAGTGCAGCTAACAGATATACATCGGTCT 531
QY 349 ProAlaAspGlnMetAspValIleGlyValGlyIleAspPheGluAspAsnIleAla 368
Db 532 CCAGCAGCTGCAAGCAAAAGTTATTACAGTGGACCGCTTCACAGTAGTGTGTATACAA 591
QY 369 ArgPheSerSerArgLysMetThrThrTrpGluLeuProGlyGlyTyrGlyArgMetLys 388
Db 592 AGCTTCTCAAGCAGAGGG-----CCAACTCAGACGCGGAGGCTTAAG 633
QY 389 ProAspIleValThrTyrGly-----AlaGlyValArgLysSerGlyValLysGly 405
Db 634 CTTGAGTGTGTGCTCCAGCAACTGGATAATTCTCCAGACAGCAAGTGGAACTAGCATG 693
QY 406 Gly-----CysArgAlaLeuSerGlyThrSerValAlaSerProVal 419
Db 694 GGTCAACCAATTAATGACTATTACACAGCAGCTCTGGGACATCAATGCAACTCCCTAC 753
QY 420 ValAlaGlyValAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnPro 439
Db 754 GTAGCTGTATTGACGCTCTTTG-----CTCCAAGCAGCACCCGAGCTGGAGCTCA 804
QY 440 AlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPhe 459
Db 805 GACAAAGTAAAAACAGCCCTCATAGAAAGTCT----- 837
QY 460 GluGlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIle----- 474
Db 838 -----GATATCGTAAGCCAGATGAATACCCGATATACCTAC 876
QY 475 -----LeuAsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAsp 490
Db 877 GTGCAGGTAGGGTTAATGCATACAAG-----GCTATTAACCTACCATAC 921
QY 491 LeuThrGluCysProTyrMetTrpProTyrCysSerGln-----ProIle 505
Db 922 TATGCAAAAGCTAGTGTTCAC-TGGATATATGTTGCCAAAGAGCGAGCCAACTCAGCAGTT 980
QY 506 TyrTyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThr 525
Db 981 CGTTAT-----TAGCGGAGCTTCGTTGCTAATCACTGCCACATATTACTG 1022
QY 526 GlyArgIleValAspLysPro-AspTrpGlnProTyrLeuProGlnAsnGlyAspAsnIle 545
Db 1023 GGACATGCCATAGCAGCTTGCATCTTTACTCTACGATCC---AATGGAACAGGT 1079
```


Qy 545 eGluValAlaPheSerTyrSerValLeuTrpProTyrSerGlyTyrLeu 562
Db 1080 TGAC-----TACTCTTACACCGC-CTACTATGATTCGAAAGGTTGGTTA 1124

RESULT 13

US-10-090-624-15

; Sequence 15, Application US/10090624

; Patent No. US20020132335A1

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

; FILE REFERENCE: TAKAKURA-6

; CURRENT APPLICATION NUMBER: US/10/090,624

; PRIOR FILING DATE: 2002-03-06

; PRIOR APPLICATION NUMBER: 09/445,472

; PRIOR FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997

; NUMBER OF SEQ ID NOS: 33

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 15

; LENGTH: 1962

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Synthetic

; US-10-090-624-15

Alignment Scores:

Pred. No.:	4,21e-22	Length:	1962
Score:	302.00	Matches:	119
Percent Similarity:	43.78%	Conservative:	64
Best Local Similarity:	28.47%	Mismatches:	150
Query Match:	5.38%	Indels:	87
DB:	12	Gaps:	19

US-09-830-837-6 (1-1052) x US-10-090-624-15 (1-1962)

Qy 186 LeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMet 205
Db 403 TTAAAGGACTGGATGAGTCTGCAGCTCAAGTTATG---GCAACTTACGTTTGGAACTTG 459
Qy 206 GlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHis 225
Db 460 GGATATGATGGTTCTTGGAAATCAACAATAGGAATATTGACACTGGAAATTGACGCTTCTCAT 519
Qy 226 ProHisPheLys---AsnValLysGluArgThrAsnTrpThrAsnGluArgThrLeuAsp 244
Db 520 CCAGATCTCAAGGAAAGTAATGGTGGGTAGATTTTGTCAATGGTAGAGTTATCCA 579
Qy 245 ---AspGlyLeuGlyHisGlyThrPheValAlaGlyVal-----IleAla 258
Db 580 TACGATGACCATGACATGCACTGCACTGATGCTCAATAGCAGCTGGTCTGGAGCAGCA 639
Qy 259 SerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPhe 278
Db 640 AGTAATGGCAAGTACAAGGAATGGCTCCAGAGGACTAAGCTGGCGGGAATTAAGGTTCTA 699
Qy 279 ThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTrpAlaIleLeu 298
Db 700 GGTCGGATGGTCTGGAACATATCTACTATTAATTAAGGAGTTGAGTGGCGGCTGAT 759
Qy 299 LysLys-----IleAspValLeuAsnLeuSerIleGly----- 309
Db 760 AACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTGTTTCAAGCCAGAGCTCA 819
Qy 310 ---GlyProAspPheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsn 328
Db 820 GATGGTACTACGCTCTTAAGTCAGGCT---GTTAATGCAGCTGGGATGCT----- 867

SEQ ID NO 5
LENGTH: 4765
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-10-090-624-5

Alignment Scores:

Prod. No.: 1,14e-19 Length: 4765
Score: 285.00 Matches: 185
Percent Similarity: 33.44% Conservative: 116
Best Local Similarity: 20.56% Mismatches: 313
Query Match: 5.07% Indels: 286
DB: 12 Gaps: 39

US-09-830-837-6 (1-1052) x US-10-090-624-5 (1-4765)

Oy 245 AspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAla----- 258
Db 1204 GATGGTCAAGTACGAGAACTCACGTAGCTGGAACCTGCTGGTTAGCAGACCAACAT 1263
Oy 258 ----- 258
Db 1264 GATGCTGGATTGGCTCAGTATGATCTGTGGTGAATGGGAAGTCTCTCAAGACTCTAT 1323
Oy 259 -----SerMetArgGluCysGlnGlyPheAlaProAspAlaGlu 271
Db 1324 GATGGATTATAGAACGTTTACACAGACACCGGTGAGGGTGTGCTCCAGGTGCCAA 1383
Oy 272 LeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAsp 291
Db 1384 ATAATGGCAATAAGAGTTCTTAGGAGTGTAGCGGGGTAGCATGTGGGATATATTAGAA 1443
Oy 292 AlaPheAsnTyrAlaIleLeuLysLysIleAspValLeuAsnLeuSerIleGlyGlyPro 311
Db 1444 GATATGACATACGACGACCAACCCATGTCACAGCTTATAAGCATGAGTCTCGGTGA--- 1500
Oy 312 AspPheMetAspHisProPheValasp-----LysValTrpGlu 324
Db 1501 -----AATGCTCCATAGTATGATGGTATGATCCAGAAAGGTTGCTGGTATGAG 1551
Oy 325 LeuThrAlaAsn---AsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyr 343
Db 1552 CTTACCGAAGTACGGTGTGTATCTGATATAGTCTGAGAAATGAAGTCTCGCAT 1611
Oy 344 GlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly----- 359
Db 1612 AACATGTTGGAGTCTGCTGTGTCACAAGCCATACCTGTGGAGCTGTCAGTG 1671
Oy 360 -----GlyIleAspPheGluasp----- 365
Db 1672 CCCATTACGTTGGAGTTTATGTTTCCCAAGCACTTGGATATCCTGATCTATGATTC 1731
Oy 366 -----AsnIleAlaArgPheSerSerArgGlyMetThr 376
Db 1732 TATTACTCCCGCTACACAACGTTAGATAGCATCTTCTCAAGCAGAGG----- 1785
Oy 377 ThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAla 396
Db 1786 -----CCGAGATAGATGTGAAATAAAACCAATGATGCTGCTCCAGGTTAC 1833
Oy 397 GlyValArgGlySerGlyValLys-----GlyGlyCysArgAlaLeuSerGlyThrSer 414
Db 1834 GGAATTTACTATCCCTCGCGATGTGGATGGCGGACCTGACTTTCATGTCGGAACCTCG 1893
Oy 415 ValAlaSerProValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArg 434
Db 1894 ATGCTATCCACATGTCACGGTGTGCTGTCATCTCTATAGCGGGCGAAGCCGAG 1953
Oy 435 GluLeu---ValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeu 453
Db 1954 GGAATATACTACAATCCAGATATATAAGAGGTTCTTGAGAGCGGTGCACCTGCTT 2013

Oy 454 ProGly-----ValAsnMetPheGluGlnGlyHisGlyLysLeu 466
Db 2014 GAGGAGATCCATATATACTGGCAGAACTACACTGAGCTTGACCAAGGTCATGGTCTTGT 2073
Oy 467 AspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAlaSerLeuSerPro 486
Db 2074 AACGTTACCAAGTCTGGGAAATCCTTAAGCT----- 2106
Oy 487 SerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSerGlnProIleTyr 506
Db 2107 -----ATAAACGCGACCACTCTCCCAATTGTTGATCCTGGCAGACAACTCC----- 2154
Oy 507 TyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGlyValThrGly 526
Db 2155 TACAGCGACTTTCGGGAGTACTTGGTGTGAGCGCTTATAAGAGGTTCTTACGAGGAAC 2214
Oy 527 ArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGlyAsnIleGlu 546
Db 2215 TCTATACCTGACATGTCGAGTGG-----CACATTAAAGTACGTAGGCGACAGGATAC 2268
Oy 547 ValAlaPheSer---TyrSerValLeuTrp-----ProTrp---SerGlyTyrLeu 562
Db 2269 AGAATCTTTGAGATCTATGCATGAGCCATGGATTAAGCTTTTGTGAGTGAAGTGA 2328
Oy 563 AlaIleSerIleSer-----ValThrLysLysAlaAlaSerTrpGluGlyIleAla 579
Db 2329 ATTCAGAGAAACATACCGAGTGTGCTTAGGTGAATATCATGTAGAGGCTCTGAG 2388
Oy 580 GlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAla 599
Db 2389 CCAGGCTCTCTATGTTGGAAGGATAATCATGATGATCCAAACAGCCAGTATTTGAGAC 2448
Oy 600 GluGlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArg 619
Db 2449 GAGATCTTGAACAATTTGTTATCCGAGAAGTTC-----ACTCCTGAGAAC 2496
Oy 620 SerLysArgValLeuTrpAspGln-----TyrHisAsnLeu 631
Db 2497 AATTACACCTCACTGTTATGATATTAATGTCAGAAATGGTGTGACTCACCACTTCTC 2556
Oy 632 ArgTyrProGly----- 636
Db 2557 ACTGTGCTGAGGAGTGGAGCTTCTACGCGATGACACATACCTACGCTACGGTCTG 2616
Oy 637 -----TyrPheProArg----- 640
Db 2617 TACAGACCATGGAATGTTGTTCCCATACCACTAGATATCTTCCGCTGCAGTC 2676
Oy 641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
Db 2677 TCAATCCATGCCCTGAAACTGGGAGCTAGTATGGACTGA-----TTTAAC 2724
Oy 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2725 TTGCAACCCCTCTATGAG-----TCGGGCTTCTTGTAGGATTTACGGAGTA 2772
Oy 681 ProPheThrCys----- 684
Db 2773 GAGATAACTCCAAAGCTTTGGTTACATTAACAGGACATACCTTGACACTACACTGAATC 2832
Oy 685 -----PheAspAlaSerGlnTyr-----GlyThrLeuLeuMetVal 696
Db 2833 TCAATTAATTAATATTTACTAATCTATGCCCCCAATTAATGCAACTCTAATCCCAAT 2892
Oy 697 Asp-----SerGluGluGluTyrPhe----- 703
Db 2893 GGCCTTGGAACTTACAATGCGAGCGTTGAAAGCGTTGGTGTGATGAGAGTCTTCTATAAG 2952
Oy 704 -----ProGluGluIleAlaLysLeuArgAspValAspAsn 716
Db 2953 GGCATTGAGTTCCTGAGGACCGCAGAGTTGAAGATTAGGATAGCAACCAAGTGT 3012

```
QY 717 -----GlyLeuSerLeuValIlePheSerAspTyrTrpAsn----- 728
Db 3013 CCGAATTCAGATCTAGACTTGTACCTTTATGACAGTAAGGCAATTTAGTGGCTTAGAT 3072
QY 729 -----Thr 729
Db 3073 GGAACCCACACAGCAGAAGAGGTGTAGTATCCTTAAGCCTGGAGTTTATTC A 3132
QY 730 SerValMetArgLysValLysPheTyrAspGluAsnThrArg----- 743
Db 3133 ATAGTAGTACATGTTTACAGCGTCACGGACGAAATGCTTAATCCACGACCAACCCTTT 3192
QY 744 -----GlnTrpTrpMetProAspThrGlyGlyAlaAsn----- 754
Db 3193 GACTTAGTGTGTTCAATGACCCCT---GATAATGGAAACATAAAGCTTGACAAAGACTCG 3249
QY 755 ---IleProAlaLeuAsnGluLeuLeuSerValTrp-----AsnMetGly 768
Db 3250 ATTATTCTTGGAGCAATGAAGCGTAGTGTGTAACCTGCAACATAACAATTGATAGAGAT 3309
QY 769 PheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAla 788
Db 3310 CATCTACAGGAGTATACTCTGTATCATAGAGATTAGAGATAATGAGGTCTACAGGAT 3369
QY 789 SerGlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThrPhe--- 807
Db 3370 ACAATACTTCAATTGCGAAATACCCATAACTTTGTTAATGACAAGCGGACTTTGCC 3429
QY 808 -----LysAspGlnGlyLeuGluValLeuLys 816
Db 3430 GTTGGTCTCACACCACAGGAGGAGTACTTGGAGAGGCTAGAAATACACTCTAATGTGA 3489
QY 817 GlnGluThrAlaValValGluAsnValPro-----IleLeuGlyLeuTyrGln 832
Db 3490 AAGCATGCCCTACACTAGAGCTGTGCCAAATGCTACAGTGTATATAGGAACACTACACC 3549
QY 833 IleProAlaGluGlyGlyArgIleValLeu---TyrGlyAspSerAsnCysLeuAsp 851
Db 3550 TACCTCACAGACGAAAAACGGTACAGTGCACATTCACGTATGCTCCAACTAAGTTCAGCAGT 3609
QY 852 AspSerHisArg-----GlnLysAspCysPheTrpLeuLeuAspAlaLeuGln 868
Db 3610 GATGAATCAGCTCATAGTATGTAAGAAGAGAACTTCAACACATAGAGAAAGACCTTCCAA 3669
QY 869 -----TyrThrSerTyrGlyValThrProProSerLeuSerHis 881
Db 3670 ATCAGAGTATCAGAGCTGAAATACTGAAGAGCACATAAATCAGCCCAAGCTTGCAATG 3729
QY 882 SerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGlu 901
Db 3730 TCATCA-----CCAGAAGCAAAATGCTTACCATAGTATCATGATGGAG 3774
```

RESULT 15

US-08-322-678-11

; Sequence 11, Application US/08322678

; Publication No. US20030077807A1

; GENERAL INFORMATION:

; APPLICANT: Gravcar, Thomas P

; APPLICANT: Bott, Richard R

; APPLICANT: Wilson, Lori J

; TITLE OF INVENTION: Subtilisin Variants

; NUMBER OF SEQUENCES: 15

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genencor International, Inc

; STREET: 180 Kimball Way

; CITY: So. San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/322,678

; FILING DATE: 13-OCT-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Horn, Margaret A.

; REGISTRATION NUMBER: 33,401

; REFERENCE/DOCKET NUMBER: GC235-2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 742-7536

; TELEFAX: (415) 742-7217

; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1140 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

US-08-322-678-11

Alignment Scores:

Pred. No.:	2,49e-20	Length:	1140
Score:	282.00	Matches:	113
Percent Similarity:	40.50%	Conservative:	64
Best Local Similarity:	25.86%	Mismatches:	166
Query Match:	5.02%	Indels:	94
DB:	7	Gaps:	17

US-09-830-837-6 (1-1052) x US-08-322-678-11 (1-1140)

```
QY 41 LeuThrLeuLysValGluPheSerSerThrVal-----ValGluTyr 54
Db 40 CTACTCATTTCTGCTGTTTCTTCTTCTCATGATCGATCGCTGCTGAAGAAGCAAGAA 99
QY 55 GluTyrIleValAlaPheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSer 74
Db 100 AAATATTATTTGCTGCTTAAAT-----GAGCAGGAAGCTGTCAGTGAGTTGTAGAA 150
QY 75 SerAlaLeuLysSerSerGluValAspAsnTrpArgIleIleProArgAsnAsnProSer 94
Db 151 CAAGTAGAGCAATACGAGGTC-----GCCATTCTCTCTGAGGA----- 192
QY 95 SerAspTyrProSerAspPheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeu 114
Db 193 -----GAGGAAGTCGAAATTTGAATTCCTTCATGAATTTGAACGATTTCCGTGT 240
QY 115 LeuThrLeuGluAspHisPro-----AsnIleLysArgValThrProGlnArgLys 131
Db 241 TTATCGTTGAGTTAAGCCCAAGAGATGTGACGCGCTTGAACCTCGATCCA----- 291
QY 132 ValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCysAsnGluThrArg 151
Db 292 -----GCGATTCTTATATTGAAGAGGATGCAGAGTA-----ACGACA 330
QY 152 TrpSerGlnLys-----TrpGlnSerSerArgProLeuArgAlaSerLeuSerLeu 169
Db 331 ATGGCGCAATCAGTCGCAATGGGCAATTAGCCGT----- 363
QY 170 GlySerGlyPheThrPheHisAlaThrGlyArgHisSerSerArgArgLeuArgAlaIle 189
Db 363 ----- 363
QY 190 ProArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGly 209
Db 364 -----GTGCAAGCCCCAGCTGCCCATACCGTGTGATTCACAGGT 402
QY 210 AlaAsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLys 229
Db 403 TCTGGTGTAAAGTTGCTGCTCCGATACAGGATATTCC---ACTCATCCAGACTTA--- 456
QY 230 AsnValLysGluArgThrAsnTrpThr---AsnGluArgThrLeuAspAspGlyLeuGly 248
Db 230 AsnValLysGluArgThrAsnTrpThr---AsnGluArgThrLeuAspAspGlyLeuGly 248
```

[illegible]

RESULT 16

```

US-08-322-678-12
; Sequence 12, Application US/08322678
; Publication No. US20030077807A1
;
; GENERAL INFORMATION:
;
; APPLICANT: Graycar, Thomas P
; APPLICANT: Bott, Richard R
; APPLICANT: Wilson, Lori J
; TITLE OF INVENTION: Subtilisin Variants
; NUMBER OF SEQUENCES: 15
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genecor International, Inc
; STREET: 180 Kimball Way
; CITY: So. San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94080
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Versio
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/322,678
; FILING DATE: 13-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

```

QY 266 PheAlaProaspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyr 285
Db 577 GTAGCCGCGAGCGGAACTATCGTGTAAAGTATAGGGCGGCGGTTCAGGCGCC 636
QY 286 ThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLeuLysIleAspValLeuAsn 305
Db 637 ATCAGCTCGATTCCCAAGATTGGAATGGGAGGGAACAATGGCATGCACGTTGTGTAAT 696
QY 306 LeuSerIleGlyProaspPheMetAspHisProPheValAspLysValTrpGluLeu 325
Db 697 TTCAGTTTAGGAAGCCCTTCG---CCAAAGTCCACACTTCAGCAAGCTGTATAGCGCG 753
QY 326 ThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThr 345
Db 754 ACTTCTAGAGCGCTTCTGTGTAGCGCATCTGGGAATTCAGGTGCA-----GGCTCA 807
QY 346 LeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyGlyIleAspPheGluAsp 365
Db 808 ATCAGGTATCCGCGCCGTTATGCGAAGCAATGGCAGTCGGAGCTACTGTACCAAAACAAC 867
QY 366 AsnIleAlaArgPheSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGly 385
Db 868 AACCGCGCAGCTTTTACAGTAGTATGCGCGAGGCTT----- 903
QY 386 ArgMetLysProaspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLysGly 405
Db 904 -----GACATTTCGCACAGCGGTGAACGTGCAGAGCACATACCCAGGTTCA 951
QY 406 GlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGlyAlaValThr 425
Db 952 ACCTATGCCAGCTTAACGGTATCATCGATGGCTACTCCTCATGTTGCGAGTGCAGCAGCC 1011
QY 426 LeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAla 445
Db 1012 CTGTGT-----AAACAAAGAACCCATCTTGTCCCAATGTACAAATCCGCAATCAT 1062
QY 446 LeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGluGlnGly 462
Db 1063 CTAAGAATACGGCAACGAGCTTAGGAAGCACCAACTGTGTATGGAAGCGGA 1113

RESULT 17

US-09-920-118-21
; Sequence 21, Application US/09920118
; Patent No. US20020102734A1
; GENERAL INFORMATION:
; APPLICANT: Menzel, Rolf
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIRECTED GENE ASSEMBLY
; FILE REFERENCE: 10424-003
; CURRENT APPLICATION NUMBER: US/09/920,118
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/222,134
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 576
; TYPE: DNA
; ORGANISM: Bacillus subtilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(576)
US-09-920-118-21

Alignment Scores:
Pred. No.: 1,34e-18 Length: 576
Score: 261.50 Matches: 63
Percent Similarity: 51.87% Conservative: 34
Best Local Similarity: 33.69% Mismatches: 77
Query Match: 4.66% Indels: 13
DB: 10 Gaps: 6

US-09-830-837-6 (1-1052) x US-09-920-118-21 (1-576)

QY 197 LeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaVal 216
Db 1 ATTAAGCGCGCGCTCTCTACTCTCAAGGCTACACAGGTTCTAACGTAAGTAGCCGTA 60
QY 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn 236
Db 61 ATTGACAGCGGAATGACTTCTCATCTCCAGCTTG---AACGTCAGAGCGGAGCAAGC 117
QY 237 TrpThrAsnGluArgThr-----LeuAspAspGlyLeuGlyHisGlyThrPheValAla 254
Db 118 TTCGTACCTTCTGAACAAACCCATCAACAGATGGCAGTTCTCAGCGCACACATGTAGCC 177
QY 255 GlyValIleAlaSerMetArgGlu-----CysGlnGlyPheAlaProAspAlaGlu 271
Db 178 GTCAGGTTGCCGCACTTAATACTCAATCGGTGTTTGGCGCTAGGCCCAACGCAATCG 237
QY 272 LeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAsp 291
Db 238 TTATATGCAGTAAGATTTCTTGATTCACAGGAACGCGCAATACAGCTGGATTATTATAC 297
QY 292 AlaPheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSerIleGlyGlyPro 311
Db 298 GGCATTGAGTGGGCCATTTCCAAACAAATGACGTGATTAAACATGACCTTGGCGGACCT 357
QY 312 AspPheMetAsp-----HisProPheValAspLysValTrpGluLeuThrAlaAsnAsn 329
Db 358 TCTGTTTACAGCTTTGAAATCAGTCGTGTATAGAGCC-----GTAGCCAGCGGT 408
QY 330 ValIleMetValSerAlaIleGlyAsnAspGlyPro-----LeuTyrGlyThrLeuAsn 347
Db 409 ATCGTGTGTTGTCGAGCGCGGAATGAAGGCACTTCCGGAAGCTCAACGACACATCGGC 468
QY 348 AsnProAlaAspGlnMetAspValIleGlyValGlyIleAspPheGluAspAsnIle 367
Db 469 TATCTCTCAAAATATCTTCTACCATTCGCGTAGTGGCTGAACACAGCACCAARAGA 528
QY 368 AlaArgPheSerSerArgGly 374
Db 529 GGTTCATTCTCAAGCGGTAGT 549

RESULT 18

US-09-927-827-33
; Sequence 33, Application US/09927827
; Publication No. US20030036176A1
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
; FILE REFERENCE: 38-10(15824)B
; CURRENT APPLICATION NUMBER: US/09/927,827
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: US 60/279,493
; PRIOR FILING DATE: 2001-03-28
; NUMBER OF SEQ ID NOS: 69
; SEQ ID NO 33
; LENGTH: 3788
; TYPE: DNA
; ORGANISM: Xanthomonas campestris
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1001)..(2788)
US-09-927-827-33

Alignment Scores:
Pred. No.: 3,38e-17 Length: 3788
Score: 260.50 Matches: 111
Percent Similarity: 43.81% Conservative: 59
Best Local Similarity: 28.61% Mismatches: 124
Query Match: 4.64% Indels: 96
DB: 9 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-927-827-33 (1-3788)


```
QY 223 uLysHisProHisPheLys-----AsnVa 231
      |||||
Db 641 GGATCATCCGGCTTGAAACAAAGTACAGAGGCTACACCCCTTAAACCGAAGCCTC 700
QY 231 lLysGluArgThrAsnTrp-----ThrAsnGluArgThrLeuAspAspGlyLe 247
      |||||
Db 701 AGACAATGAGTTCAACTGGTTTGCATCGGTAAAGCAATAAGAAACCCCATATGATGAGTT 760
QY 247 uGlyHisGlyThrPheValalaGlyValle-----AlaSerMetArgGluCy 263
      |||||
Db 761 GGCCACGACGACATGTCACAGGACGATGCTCGGCTCAGAGCGCGGCGCAAAACCA 820
QY 263 sGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThr----- 279
      |||||
Db 821 GATCGGTGTCGCCCGACGCGGAATGATAGTGTCAAGGCCTTTTCAGAGGACGCGG 880
QY 280 -----AsnAsnGlnValSerTyrThrSerTyrPheLeu-----AspAlaPheAs 294
      |||||
Db 881 AGATGAAATCCCTTTGGCTCCAGGAGAAATGGATTGGCCCGCGAAAGATCGAAAGG 940
QY 294 nTyrAlaIleLeuLysIle-----AspValLeuAsnLeuSerIleGlyGlyProAspPh 313
      |||||
Db 941 AAAAGCGACCCCTGAAAGGCTCCCGATGCTGCACAACTCTTGGCAGGGAAGCGCG 1000
QY 313 eMetAspHisProPheValaspLysValTrpGluLeuThrAlaAsnValIleMetVa 333
      |||||
Db 1001 ACTCGATGATGTTATTTGATATTGTCAAAGCTTGGCGGCGCATGATCTTCCCTGC 1060
QY 333 lSerAlaIleGlyAsn-----AspGlyProLeuTyrGlyThrLeuAs 347
      |||||
Db 1061 ATTGACGGCGGCAATGTCAGCGAATTCGAACGCGCGACCG-----GGTTCAGTTGA 1114
QY 347 nAsnProAlaAspGlnMetAspValIleGlyValGlyGlyIleAspPheGluAspAsnI 367
      |||||
Db 1115 AAACCGCTCAACTATCCGCAATCTTCGGATGCGGAGTTGACAGCTCAACCGCGTT 1174
QY 367 eAlaArgPheSerSerArgGlyMetThrThrTrpGluLeuProGlyGlyTyrGlyArgMe 387
      |||||
Db 1175 AGCGGACTTTTCATTCAAGG-----CCTTCGCCATATGACGAAAT 1216
QY 387 tLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGlyValLysGlyGlyCy 407
      |||||
Db 1217 CAAGCCGATATTTCTGCACCGTGCATCGCTCCGCTGTATCCGGGACATAAGTA 1276
QY 407 sArgAlaLeuSerGlyThrSerValAlaSerProValAlaGlyAlaValThrLeuLe 427
      |||||
Db 1277 TGCTGAATGAAGGAGATCGATGCGACCGCATGTATCGGCGATGCTGCCCTTAAT 1336
QY 427 uValSerThrValGlnLysArgGluLeuValAsnPro-----AlaSerMetLy 443
      |||||
Db 1337 G-----AGGAA-----GCCAATCCGGACCTCACAGTCGATGAATCGA 1375
QY 443 sGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGluGln----- 461
      |||||
Db 1376 GGCATCTCTGCTGAAACCGCGACGCGCTCTGATGCAAAACGTTTAAAAAATCCCGAA 1435
QY 462 -----GlyHisGlyLysLeuAspLeuLeuArgAlaTyr 472
      |||||
Db 1436 CAACGCTACGGATACGGTTGGTCAACGCTTAAAGCTTTT 1478

RESULT 20
US-09-783-590-349
; Sequence 349, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2C1
; CURRENT APPLICATION NUMBER: US/09/783,590
```

```
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 349
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (51)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (82)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (205)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (206)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (210)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (230)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (231)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (263)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (298)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (308)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (309)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (310)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (314)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (349)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (355)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (368)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (380)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (381)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

LOCATION: (411)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (421)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (443)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (447)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (448)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (455)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (472)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (480)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (486)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (489)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-349

Alignment Scores:
Pred. No.: 3,81e-17 Length: 497
Score: 247.00 Matches: 50
Percent Similarity: 76.47% Conservative: 2
Best Local Similarity: 73.53% Mismatches: 8
Query Match: 4.40% Indels: 8
DB: 10 Gaps: 2

US-09-830-837-6 (1-1052) x US-09-783-590-349 (1-497)

QY 735 ValLysPheTyArgGluAsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsn 754
DB 2 ATTCGCCAAGTGACCACTTCCAGGAGTGGTGGATCCGGATACGAGGCTAAC 61
QY 755 IleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTrp 774
DB 62 ATCCAGCTCTGAATGAGCTGCTGTGTGGAACTGGGTTCAGCAIGGCCCTGAT 121
QY 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyTrpAlaSerGlyCysSerIleAla 794
DB 122 GAAGGGAGTTACCTCGCCCAACCATGAC-----AGTAAGGCT 160
QY 795 LysPheProGluAspGlyValVal 802
DB 161 CTGTTTCTGTAG---GGCGTGGTC 181

RESULT 21

US-09-974-300-1934
Sequence 1934, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
TITLE OF INVENTION: Methods For Monitoring Multiple Gene
FILE REFERENCE: 10085-500-US
CURRENT APPLICATION NUMBER: US/09/974,300
PRIOR FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/660,598

PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1934
LENGTH: 1329
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-1934

Alignment Scores:
Pred. No.: 7,04e-16 Length: 1329
Score: 241.50 Matches: 89
Percent Similarity: 45.19% Conservative: 52
Best Local Similarity: 28.53% Mismatches: 108
Query Match: 4.30% Indels: 64
DB: 10 Gaps: 13

US-09-830-837-6 (1-1052) x US-09-974-300-1934 (1-1329)

QY 179 ArgHisSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGln 198
DB 368 CCTCTCTGACGTAGCGGTGAATCAAGCCATGCCAAAGAGTGTGCAGAAACAAT-CAG 426
QY 199 AlaAspValLeuTrpGlnMetClyTyThrGlyAlaAsnValArgValAlaValPheAsp 218
DB 427 ACG-----CTGACAGGCAAGGAGTGCACAGTCGCTGTCATTGAT 465
QY 219 ThrGlyLeuSerGluLysHisProHisPheLys-----AsnValLysGluArgThr 235
DB 466 ACGGCGCTA-----TACCTCCAGAGATCTTGAAGGAGGATCAGGCGCTTTCAA 516
QY 236 AsnTrpThrAsnGluArgThr-----LeuAspAspGlyLeuGlyHisGlyThrPheVal 253
DB 517 GACTTATTCACACAGAGAACACCTATGATGACAAT---GGCAGCGGCACACACTGC 573
QY 254 AlaGly-----ValIleAlaSerMetArgGluCysGlnGlyPheAlaPro 268
DB 574 GCCGCTGATCCTTCGCGAAACGAGCGGCTCATCGGGTCAGTACCGCGGACCTCTCT 633
QY 269 AspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyThrSerTrp 288
DB 634 GAAGCAGAACTTCGCGTGTAAAGTATTGGACAAAATGGGATCCGGATCCGCTCGAAC 693
QY 289 PheLeuAspAlaPheAsnTyTrpAlaIleLeu-----LysLysIle 301
DB 694 GTCAATTCAGGCGTAGATTGGTGCATTCAATTCAATAAGGAAAATCTGTATGATCCGATC 753
QY 302 AspValLeuAsnLeuSerIleGlyGlyProAspPheMet-----AspHis 316
DB 754 GACATATTCATGTCATTTGGTGCACAGACCTTCGCTACGAGATGAAGAAGAAT 813
QY 317 PropheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIle 336
DB 814 CCAGTCGTTAANGCTGTTCATGCAGCATGGGACCGAGGCAATGTTGTATGTGGCGACCC 873
QY 337 GlyAsnAspGlyProLeuTyTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle 356
DB 874 GGCAACTCCGGTCTCTGATCGCAACAGATTGCCAGCCGGGTCTCAGCAGCAAGATATT 933
QY 357 GlyValGlyGlyIleAsp-----PheGluAspAsnIleAlaArg 369
DB 934 ACAGTCGGAGCGCTTGGATGACAGGGATACAGTCACCGCGGAGGATGACGATGCGCTCT 993
QY 370 PheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyTrpGlyArgMetLysPro 389
DB 994 TATTCAGCAGAGGCGGCAATC-----TATGGTCAAGTCAAAACCG 1035
QY 390 AspIleValThrTyTrpGlyAlaGlyVal-----ArgGlySerGly----- 402
DB 1036 GACTTGTGGTGGTACCGGCGACAAATATTACGTGCTGCTTCCACCGGATCTTTCTCGAT 1095


```
QY 403 -----VallySGlyGlyCysArgAlaLeuSerGlyThrSerVal 415
Db 1096 AAGCTGCAAAAACAAACAGAGTCGGCAAAATATATACATGTGTCGGAACCTCGATG 1155
QY 416 AlaSerProValValAlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGlu 435
Db 1156 GCTACGCCGCTGTCGCGAGGAATGCGGCACCTATC-----CTTCAGCAAGCCCG 1206
QY 436 LeuValAsnProAlaSerMetLysGlnAlaLeuIle 447
Db 1207 GCACAGAACCTGATGATCAAGTCAACACAGCTGCTAATG 1242

RESULT 22
US-09-966-921A-5
; Sequence 5, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097-200-US
; CURRENT APPLICATION NUMBER: US/09/966, 921A
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1330
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-966-921A-5

Alignment Scores:
Pred. No.: 1,37e-14 Length: 1330
Score: 229.50 Matches: 109
Percent Similarity: 39.46% Conservative: 65
Best Local Similarity: 24.72% Mismatches: 160
Query Match: 4.09% Indels: 107
DB: 10 Gaps: 19

US-09-830-837-6 (1-1052) x US-09-966-921A-5 (1-1330)
QY 57 IleValAlaPheAsnGlyThrPheThrAlaLysAlaArgAsnSerPheIleSerSerAla 76
Db 22 ATTGTGCGAAGCAGCCGACTACTATCTTCTGTGCTTTAGTTCATCGATCGATCGGCT 81
QY 77 -----LeuLysSerSerGluValAspAsnTrpArgIleIleProArg 90
Db 82 GCACCTGCAAAAGACAAAGTTGAGGTAAAGAACAAAGATTATCATCGTGTGCTAATCAAA 141
QY 91 -----AsnAsnProSerSerAspTrpProSerAspPheGluVal----- 103
Db 142 GCACCACTACATCAATCAGTACTTTTCAATCAATCAGATGATGTCGGTTGGGATTTGGC 201
QY 104 -----IleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGlu 118
Db 202 AAAGAGGGATTACACACAGATGTTGATGCCAACAG-----CTCCAAACGCTTCAA 252
QY 119 AspHisProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr 138
Db 253 AGCAACAAAGACATTCAAATTCAGAGGTAAATGAAATGACAGTAGAAGTCTGTACAA 312
QY 139 AlaGluSerAspProThr-----ValProCysAsnGluThrArgTrpSerGlnTrpGln 157
Db 313 GAAAAGCGGAGTAGTCGCGCGGTACCAAGTACACAAACCCCTGGGGCATAAAG----- 366
QY 158 SerSerArgProLeuArgAlaSerLeuSerLeuLysSerGlyPheTrpHisAlaThr 177
Db 367 -----TCAATTTTATATGATCAATCAATCAATCAATCAATCAATCAATCAATCA 402
QY 178 GlyArgHisSerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeu 197
```

```
Db 403 GGA----- 405
QY 198 GlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPhe 217
Db 406 -----GGCAGCGGAATTAAGTAGTACTGTTTAA 432
QY 218 AspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArg---ThrAsn 236
Db 433 GATACAGGGGTTTATACAGCCATTAGATTAGCTGGTCTGCCGAGCAATCCAGGAT 492
QY 237 TrpThrAsnGluArgThrLeuAspAspGly-----LeuGlyHisGlyThr 251
Db 493 TTTACCCCAATCTAATCCTTTTAGTAGATGTTTCATGCACCGATGCCCAAGGCGATGTACA 552
QY 252 PheValAlaGlyValIle-----AlaSerMetArgLysGlnGlyPhe 266
Db 553 CATGTTGCCGGAAGTGTATTGGCGCATGGAGCGCAAGTAAATGGACAGCGGTTTACGGGGTG 612
QY 267 AlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAsnGlnValSerTyrThr 286
Db 613 GTCCCGCAAGCCAACTATGGCATATAAAGTATTAGGAGATAACGCCAGCGGATCTCT 672
QY 287 SerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLysIleAsp----- 302
Db 673 -----GATGATATTGCAGCAGCTATCAGACATGTAGCTGATGAAGCTTCACGT 720
QY 303 -----ValLeuAsnLeuSerIleGly-----GlyProAspPheMetAsp 315
Db 721 ACAGGTTCCAAAGTAGTAAATTAATATGTCGTAGTTCATCTGCCAAGGATCATTTGATT 780
QY 316 HisProPheValAspLysValTrpGluLeuThrAlaAsnValIleMetValSerAla 335
Db 781 GCTAGTGCAGTAGATTATCATAT-----GGAAAAGGTGTATTATCTGTTGTCGG 831
QY 336 IleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspVal 355
Db 832 GCTGTTAATAGTGGTCCAGCAGCAATAATCGGCTTCTCGCGGGCTTGTAAATGCA 891
QY 356 IleGlyValGlyGlyIleAsp---PheGluAspAsn-----IleAlaArgPhe 370
Db 892 GTGGCAGTAGCGGCATTGGAGAAATGTCAGCAAAATGGAACTATTCGATGCTGATTC 951
QY 371 SerSerArgGlyMetThrThrTrpGluLeuProGlyGlyTyrGlyArgMetLysProAsp 390
Db 952 TCATCTAGAGGAATCCGCA-----ACTGCTGGAGATTATATCATTCATCAAGACGCTGAT 1005
QY 391 IleValThrTyr-----GlyAlaGlyValArgGlySerGlyValLysGlyCysArg 408
Db 1006 ATTGAAGTTTCAGCTCCGGGAGCAGTGTAGAGTCTACATGTGTACACTGCGCGTTAAT 1065
QY 409 AlaLeuSerGlyThrSerValAlaSerProValValAlaGly----- 422
Db 1066 ACGATCAGCGGTACATCAATGGCTACACCTCATAGCTAGCTAGCTGCTGCTAAAATCTGG 1125
QY 423 ---AlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSer 441
Db 1126 TCAGCGAATACCTTCATTAACTCATGCCCAACTGCCGACAGAAATTCGCAAAATCCGCTAAA 1185
QY 442 Met 442
Db 1186 GTA 1188

RESULT 23
US-09-966-921A-1
; Sequence 1, Application US/09966921A
; Patent No. US20020103100A1
; GENERAL INFORMATION:
; APPLICANT: Outtrup, Helle
; APPLICANT: Lassen, Soren
; APPLICANT: Pedersen, Poul
; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides Having Proteolytic Activity
; FILE REFERENCE: 10097-200-US
```

```

; CURRENT APPLICATION NUMBER: US/09/966, 921A
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1306
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (50)..(1303)
; OTHER INFORMATION:
; NAME/KEY: mat_peptide
; LOCATION: (371)..( )
; OTHER INFORMATION:
US-09-966-921A-1

Alignment Scores:
Pred. No.: 1,92e-14 Length: 1306
Score: 228.00 Matches: 114
Percent Similarity: 38.76% Conservative: 67
Best local Similarity: 24.41% Mismatches: 161
Query Match: 4.06% Indels: 125
Gaps: 21

US-09-830-837-6 (1-1052) x US-09-966-921A-1 (1-1306)
QY 26 GluLysSerPheGluLysAlaProCysProGlyCysSerHisLeuThrLeuLysVal 45
DB 55 GAAAGAGAGCATT-----TGCAGCCACATT----- 84
QY 46 GluPheSerThrValValGluTyrGluTyrIleVal-AlaPheAsnGlyTyrPheTh 65
DB 85 -----CAGTATTAGGATGCGATTATCCGTTATTTCAACAGGAGCACT 126
QY 65 rAlaLysAlaArgAsnSerPheLeuSerAlaLeuLysSerSerGluValAlaAsnTr 85
DB 127 TCAAAAGACAAA-----GTTGAGGTGAAGAACAGATTCATA 165
QY 85 pArgIleleProArg-----AsnAsnProSerAspTyrProSerAspPheGluVa 103
DB 166 TCGTGTGTAATCAAGACCACTACATCAATCACTACTTTTCAATCAATACAGATGT 225
QY 103 l-----lleglnlleLysGluLysGlnLysAlaGl 113
DB 226 CGTTGGGNTTTGGCAAGAGGGATTACACAGATGTGTGTCCTCAACAG----- 277
QY 113 yLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysValPh 133
DB 278 -CTCAACGCTTCAAGCAACAAAGACATTCAATTCAGAGTAATCAATCAATGACACT 336
QY 133 eArgSerLeuLysTyrAlaGluSerAspProThr-----ValProCysAsnGluThrArgTr 152
DB 337 AGAACTCTTACACAGAAAGGGGAGTACGCGGTACCAAGTACACAAACCCCTTG 396
QY 152 pSerGlnLysTrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGl 172
DB 397 GGCATAAAG-----TCATTTTATTAATGATCAATC 426
QY 172 yPheTrpHisAlaThrGlyArgHisSerSerArgLeuLeuArgAlaIleProArgGl 192
DB 427 AATTCAAAACAACTGGA----- 445
QY 192 nValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnVa 212
DB 446 -----GGCAGCGGAAT 456
QY 212 lArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnVally 232
DB 457 TANGTACTCTTTTAGATACAGGGGTTTATACAACCATTTAGATTACTGGTCTGCG 516
QY 232 sGluArg---ThrAsnTrpThrAsnGluArgThrLeuAspAspGly----- 246
DB 246 ----- 246

```

```

517 CGACAAATGCAAGGATTTTACCCTAATCTTATAGTAGTGGTTCATGCACCGATCG 576
247 -LeuGlyHisGlyThrPheValAlaGlyValIle-----AlaSerMetar 261
577 CCAAGGCGATGGTACACATGTTCGCGAAGTGTTCGCGCATGGAGCGACGTAAAGACA 636
261 gLUCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnAs 281
637 AGCGTTTACGGGTGGTCCGCAAGCAACTATGGCATATATAAGTATTAGGAGATAA 696
281 nGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLysLys 301
697 CGCAGCGGATCTCT-----GATGATATTGCAGCAGCTATCAGACATGATG 744
301 eAsp-----ValLeuAsnLeuSerIleGly-----G1 310
745 TGATGAAGCTTTCACGTACAGGTTCACAAAGTAGTAATTAATATGCTAGTTCATCTGC 804
310 yProAspPheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnAsnVa 330
805 CAAGGATTCATTGATTGCTAGTGCAGTAGATTATGATAT-----GGAAAAGGTGT 855
330 lIleMetValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAl 350
856 ATTATCGTTCGCTCGCGTGTGTAATAGTGGTCAGGACGACAAATCAATCGCTTCCTGG 915
350 aAspGlnMetAspValIleGlyValGlyIleAsp---PheGluAspAsn----- 366
916 CGGCTGTGTAATGTCAGTGCAGTACGCGCATGGAGATGTCAGCAAAATGGAACCTTA 975
367 -----lIleAlaArgPheSerSerArgGlyMetThrTrpGluLeuProGlyGlyTyrGl 385
976 TCGAGTACGCGATTCTCATCTAGAGGAATCCGCA-----ACTGCTGGAGATTATAT 1029
385 yArgMetLysProAspIleValThrTyr-----GlyAlaGlyValArgGlySerGlyVa 403
1030 CATTCAGAGCGGTGATTTAGAGTTTCAGCTCCGCGAGCAAGTAGTAGTCTCATCGTA 1089
403 lLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValAlaIleGly-- 422
1090 CACTGCGGTATATAATACATCAGCGGTACATCAATGCTACACCTCATGAGTGGT 1149
423 -----AlaValThrLeuValSerThrValGlnLysArgGluLe 436
1150 AGCTGCTAAAATCTGTCAGCGCAATCTTCAATTAAGTCATAGCAACTGCGCACAGAATT 1209
436 uValAsnProAlaSerMet 442
1210 GCATAATCGCGCTAAGTA 1228

```

```

RESULT 24
US-10-209-812-1
; Sequence 1, Application US/10209812
; Publication No. US20030087785A1
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Osten, Claus von der
; APPLICANT: Andersen, Kim Vilbourn
; APPLICANT: Ernst, Steffen
; APPLICANT: Roggen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; FILE REFERENCE: 5666.200-US
; CURRENT APPLICATION NUMBER: US/10/209,812
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: US/09/417,359
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: PA 1998 01301
; PRIOR FILING DATE: 1998-10-13
; PRIOR APPLICATION NUMBER: PA 1999 01418
; PRIOR FILING DATE: 1999-10-04
; PRIOR APPLICATION NUMBER: 60/105,624
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/157,426

```

;; PRIOR FILING DATE: 1999-10-04

;; NUMBER OF SEQ ID NOS: 6

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1

;; LENGTH: 840

;; TYPE: DNA

;; ORGANISM: Bacillus sp.

;; FEATURE:

;; NAME/KEY: CDS

;; LOCATION: (1)...(840)

US-10-209-812-1

Alignment Scores:

Pred. No.: 1,08e-14 Length: 840

Score: 227.50 Matches: 85

Percent Similarity: 41.24% Conservative: 35

Best Local Similarity: 29.21% Mismatches: 128

Query Match: 4.05% Indels: 43

DB: 9 Gaps: 13

US-09-830-837-6 (1-1052) x US-10-209-812-1 (1-840)

QY 195 GlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAlaAsnValArgVal 214

DB 46 CAAACACCTCAACCCCTGCTGGGATGTAACCCGTGAAGCAGCACTCAACCGTG 105

QY 215 AlaValPheAspThrGlyLeuSerGluLysHisProHisPhe-----LysAsnValLys 232

DB 106 GCGGCTTGATTCGGGATGGATTAACCAACCCCTGATCTTGCAAGAAAGTAATAAAA 165

QY 233 -----GluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeu 247

DB 166 GGTACGACTTATCGACAGGGACAAT-----AACCAATGATCTTAAC----- 210

QY 248 GlyHisGlyThrPheValAlaGlyValIleAlaSerMetArgGlu-----Cys 263

DB 211 GGATGTTGGTACCATCTGTCGGTACTGTGCTGATACGAAATGGAATGGCGTA 270

QY 264 GlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsnGlnVal 283

DB 271 GCGGTATGACACAGATACGAGATCCCTGGGTACGGGTCTTGTATGCCAATGAAGT 330

QY 284 SerTyThrSerTrpPheLeuAspAlaPheAsnTyAlaIleLeuLysLysIleAspVal 303

DB 331 GGTCTACTTGACAGCATTCCTCAGGTATCCGCTATGCTGCTATCAAGGGCAAGGTA 390

QY 304 LeuAsnLeuSerIleGlyGlyPro-----AspPheMetAspHisProPheValAspLys 321

DB 391 CTCAACCTCTCCCTGTTGCGAATGCAACTCCACAACTCTTAAGAGTCCGCTCGACTAT 450

QY 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341

DB 451 GCATGGAAC-----AAAGAGCTGTAGTGTGCTGCTGCGAGGAATGACAAT--- 498

QY 342 LeuTyThrGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyValGlyIle 361

DB 499 ---GTATCCCGTACATTCCAACAGCTTCTTACCCTAATGCCATTCCAGTAGTGCAT 555

QY 362 AspPheGluAspAsnIleAlaArgPheSerArgGlyMetThrTrpGluLeuPro 381

DB 556 GACTCCAAATGATCGAAAGCATCTCTCCAATTACGGA-----ACGTGG----- 600

QY 382 GlyGlyTyThrArgMetLysProAspIleValThrTyThrGlyAlaGlyValArgGlySer 401

DB 601 -----GTGGATGTCATCTGCTCCAGGTGTGAACATAGCATCAACC 639

QY 402 GlyValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAla 421

DB 640 GTTCCGAATAATGGCTACTCTACATGCTGTGATGCTCCATGCCATCCCTCACGTGGCC 699

QY 422 GlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSer 441

DB 700 GGTGTTGGCTGCTTTGTGGCAAGTCAAGGTAAAG-----AATAACGTACAA 744

QY 442 MetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMet---PheGlu 460

DB 745 ATCCGCCAGGCCATTGAGCAAAACCCGCGATAGATCTCTGGCACTGGAAACAACTTCAG 804

QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAla 471

DB 805 -----TATGGTAAATCAACTCAACAAAGACT 831

RESULT 25

US-09-974-300-1933

;; Sequence 1933, Application US/09974300

;; Patent No. US20020146721A1

;; GENERAL INFORMATION:

;; APPLICANT: Berka, Randy M.

;; TITLE OF INVENTION: Methods for Monitoring Multiple Gene

;; TITLE OF INVENTION: Expression

;; FILE REFERENCE: 10085.500-US

;; CURRENT APPLICATION NUMBER: US/09/974,300

;; PRIOR FILING DATE: 2001-10-05

;; PRIOR APPLICATION NUMBER: 09/680,598

;; PRIOR FILING DATE: 2000-10-06

;; PRIOR APPLICATION NUMBER: 60/279,526

;; PRIOR FILING DATE: 2001-03-27

;; NUMBER OF SEQ ID NOS: 8481

;; SOFTWARE: FastSeq for Windows Version 4.0

;; SEQ ID NO 1933

;; LENGTH: 1397

;; TYPE: DNA

;; ORGANISM: Bacillus licheniformis

US-09-974-300-1933

Alignment Scores:

Pred. No.: 3.97e-14 Length: 1397

Score: 225.50 Matches: 108

Percent Similarity: 41.22% Conservative: 61

Best Local Similarity: 26.34% Mismatches: 173

Query Match: 4.01% Indels: 71

DB: 10 Gaps: 17

US-09-830-837-6 (1-1052) x US-09-974-300-1933 (1-1397)

QY 50 ThrValValGluTyThrGluTyIleValAlaPheAsnGlyTyThrPheThrAlaLysAlaArg 69

DB 10 AGTATCGTGGCTAT---TTTGTATGGCTTTTATCTATATT---GCCTTA 57

QY 70 AsnSerPheIleSerSerAla-----LeuLysSerSerGluVal 82

DB 58 TCCACATCTCTACCCGAGTGCGCAACTTCGTTCCCGATATAAAGTCCGCTGAGCTT 117

QY 83 AspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyThrProSerAspPheGlu 102

DB 118 GAGAAAGCTGAAATCATACGGTGTATGTGTCGCTGTGATAAACAACGACGCGTT--- 174

QY 103 ValIleGlnIleLysGluLysGlnLysAla---GlyLeuLeuThrLeuGluAspHisPro 121

DB 175 ATCGTGGAACTGAAAGAAAGTCCGTTGCCAAGCAAAACGCGGAGGAGAAACAAACG 234

QY 122 AsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyAlaGluSer 141

DB 235 AAGGCTTCTCTAAACACAGCTCGAAGCAAAAGCGCTGAAACACTTAAAAA-AGCGAAGT 293

QY 142 AspProThrVal---ProCys-----AsnGluThrArgTrpSerGlnLys 155

DB 294 AAACCCGAATACGACCGTGTATTTTCGGGCTTTTCTATGAAACTGCCGCCAGTG-AAA 352

QY 156 TrpGlnSerSerArgProLeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHis 175

DB 353 TTCCAAGCTGTCGCCGTCAGGAAGTTAAACGGGTTTAAACGGGTTTATCCAAACGCAACTTACAAAC 412

QY 176 AlaThrGlyArgHisSerSerArgLeuLeuArgAlaIlePro----- 190

Db 413 CTGACGGCTAAAGAAAGACGCTGACACTTCGACGACGCGCATTTATCCACAGATGG 472
 QY 191 ArgGlnValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyThrGlyAla 210
 Db 473 ATAAAGTGCCTGTC-ATCGGAGCGGATCAGGCATGGAATCGGCTATACCGGTAAA 531
 QY 211 AsnValArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsn 230
 Db 532 GGCATTAAAGTAGCGTTATTGATACGGCGTGATACACCCATCCGACTTAAGAAT 591
 QY 231 -----ValLysGluArgThrAsnTrpThrAsnGluArg 241
 Db 592 AATTCGGCCCATATAAGGCGTATGACTTTGTGATAATGATTATGATCCACAGGAACA 651
 QY 242 ThrLeuAspAspGlyLeuGly-----HisGlyThrPheValAlaGlyValIle 257
 Db 652 CCGACAGCGCGCCGCGCGGAGCGACTGATCAGCGCACTCATGTTGCAGGAACAATC 711
 QY 258 AlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgVal 277
 Db 712 GCCGGAACGACAGATTAAAGCGGTGGCGCTGAAGCAACGCTTCTTGCTACCGCGTG 771
 QY 278 PheThrAsnAsnGlnValSerTyThrSerTrpPheLeuAspAlaPheAsnTyralIle 297
 Db 772 CTCGGCCCCGGGCTCAGGCACACACGAAATGCTGTCGGGCATTGAAAAGCCGTT 831
 QY 298 LeuLysLysIleAspValLeuAsnLeuSerIleGly-----GlyProAspPhe 313
 Db 832 GCAGACGGAGCAAAAGTAGAACCTCTCATTTGGAAATTCCTGAAACAGCCCTGATTAT 891
 QY 314 MetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetVal 333
 Db 892 GCACAAACGATCGCACTGGAC-----TGGCGCATG--GCTGAAGGGGTGTCGCCGTT 942
 QY 334 SerAlaIleGlyAsnAspGlyProLeuTyrglyThrLeuAsnAsnProAlaAspGlnMet 353
 Db 943 ACGTCAACCGTAAACAGCGACCGGAAACTGGACGCTCGGTCGCCGGGAACATCAAGG 1002
 QY 354 AspValIleGlyValGlyIle-----AspPheGluAspAsnIleAla 368
 Db 1003 GTTGGCATTTCCGTCGCGCATCACAGCTCCGATATAATGATATCGGTGACGCTTCT 1062
 QY 369 ArgPheSerSerArgGlyMetThrThrTrpGluLeuProGlyGlyTyrglyArgMetLys 388
 Db 1063 TCGTATTATCATCAGCAAGGTGATG-----GGATACCAAGAGAGAAAAG 1104
 QY 389 Pro-----AspIleValThrTyrglyValGly----- 397
 Db 1105 GATCTTGAAGCATTAACGGTCAACAGGTGGAACTCGTTGAAGCGGCGCTCGGCAAGCT 1164
 QY 398 -----ValArgGlySerGlyValLysGly 405
 Db 1165 GACGATTCTCAGGCAAAAGATGTGAAGGA 1194

Search completed: June 2, 2003, 06:02:21
 Job time : 1152 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 02:29:02 ; Search time 6516 seconds

(without alignments)

4698.611 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MKLVNTWLLVLLVLCGRK.....PRVKRQLMQVHPKTPSV 1052

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09830837/runat_23052003_181924_8065/app_query.fasta_1.1223
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=ptt -THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptt -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837.ecgn_1.1.4449.eurinat_23052003_181924_8065 -NCPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5617	100.0	4338	6	AX024613 Sequence
2	5617	100.0	4338	9	HUMKTAAL
3	5487	97.7	4198	10	AF078105 Cricetulu
4	5476	97.5	3895	6	AX024609 Sequence
5	5476	97.5	3895	10	AF094821
6	5427	96.6	3788	6	AX024611
7	5427	96.6	3788	10	AF094820
8	5412.5	96.4	4204	10	BC011533
9	3069	54.6	3026	9	BC026330
10	3063	54.5	3085	9	AK091212
11	2383.5	42.4	3138	3	AF441758
12	2219.5	39.5	174920	3	AC010701
13	2219.5	39.5	294308	3	AE003595
14	1965.5	35.0	100592	8	T29J13
15	1772	31.6	115884	2	AC017581
16	1716	30.6	1650	9	HSB01435
17	1355	24.1	170371	8	AB026295
18	1208.5	21.5	105815	8	AP002536
19	850.5	15.1	113320	2	AC010662
20	812	14.5	3333	9	AK054744
21	774.5	13.8	129955	2	AC010056
22	774.5	13.8	129955	2	AC011909
23	728	13.0	214287	2	AC040169
24	567	10.1	2905	9	AB047598
25	544.5	9.7	195287	2	AC112476
26	443.5	7.9	252539	2	AC098090
27	381.5	6.8	32704	1	SC8A11
28	373.5	6.6	42527	1	SC51A
29	363.5	6.5	3348	1	D83672
30	352.5	6.3	36102	3	LMFLA325
31	342	6.1	2494	1	BACAPRJ
32	341.5	6.1	1527	1	BACSAS
33	341	6.1	1220	6	I01972
34	341	6.1	1220	6	I08532
35	341	6.1	1524	6	I03244
36	339	6.0	1473	1	BACSUBTI
37	339	6.0	1473	1	S51909
38	339	6.0	2044	6	E01654
39	339	6.0	2044	6	E02803
40	338.5	6.0	1146	1	AF368283
41	338.5	6.0	1500	1	BACSEBTL
42	338.5	6.0	1500	6	E03694
43	338.5	6.0	1500	6	I05124
44	338.5	6.0	1500	6	I07865
45	338.5	6.0	1500	6	I95503

ALIGNMENTS

RESULT 1

QY	441	SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu	460	QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThraLa	820
DB	1817	AGTATGAACGAGCCCTGATCGCGTCAGCCCGAGGCTCCCGGGGTCAACATGTTGAG	1876	DB	2897	GTCGTGATAACACAGACTTTCAAGGACCAAGGATGGAGGTTTAAAGCAGGAACAGCA	2956
QY	461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480	QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
DB	1877	CAAGGCCAGCGAAGCTCATCTGCTCAGAGCTATCAGATCCTCAACAGCTACAAGCCA	1936	DB	2957	GTTGTTGAAAGCTCCCACTTTTGGGACTTTATCAGATTCCAGTCCAGGTGGAGCCGG	3016
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr	500	QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
DB	1937	CAGGCAAGTTTGAGCCCAAGCTACATAGATCTGACTGCTCCCTACATGTGGCCCTAC	1996	DB	3017	ATTGTACTATATGGGACTCCCAATTGCTTGGATGACAGTACCAGCAGAGGACTGCTT	3076
QY	501	CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn	520	QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSer	880
DB	1997	TGCTCCCAACCCATCTACTATGGAGGAATGCCAGACGTGTATGTCACCAATCTCAAC	2056	DB	3077	TGGCTTCTGGATCCCTCCCTCCAGTACATCGTATGGGTGACACCGCTAGCCTCAGT	3136
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540	QY	881	HisSerGlyAsnArgGlnArgProSerGlyAlaGlySerValThrProGluArgMet	900
DB	2057	GGCATGGGAGTCACAGGAAGAATTGTAGATAAGCTGACTGGCAGCCCTATTGTGCCACAG	2116	DB	3137	CACCTCTGGGAACCGCCAGCGCCCTCCAGTGGAGCAGGTCTCAGTCTACTCCAGAGAGATG	3196
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGly	560	QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
DB	2117	AACGGAGACAACATTGAAGTTCCCTCTCTACTCTCTCGGTCTTATGGCCTTGTGGGCG	2176	DB	3197	GAAGGAACCACTTCTCATCGTACTCCAAGGTCTCGAGGCCCATTTGGGAGACCCAAA	3256
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580	QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
DB	2177	TACCTGGCCATCTCCATTTCTGTGACCAAGAAGCGGCTTCTGGGAAGCAATGCTCAG	2236	DB	3257	CCTCGGCCCTACCAGCCTGTCACGCTTGTCTTGGGCCNAGCCACAGCCTTTAAACGAG	3316
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600	QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
DB	2237	GGCCATGTGATGATCAGTGGCTTCCCGCAGCAGACAGACAGTCAAAAAATGCTGCAGAA	2296	DB	3317	ACGGGCCCACTAAACCTTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGGTG	3376
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSer	620	QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
DB	2297	CAGACTTCAACAGTAAGCTCCCAATTAAGGTGAAGATAATCTACTCCCGCGGAAGC	2356	DB	3377	GTGTTACCACTTTCGATCGAATCCCTCAAGTGAAGCCCTTGTCCTCGGAGAGAGC	3436
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640	QY	981	GlyAlaTrpAspIleProGlyIleMetProGlyArgTyrAsnGlnValGlyGln	1000
DB	2357	AAGAGAGTTCTCTGGGATCAGTACCACACCTCCCGTATCCACCTGGCTATTTCGCCAGG	2416	DB	3437	GCGCCTGGAGATCTCTGGAGGATCATCGCTGGCGGTACAAACAGAGGTGGGCCAG	3496
QY	641	AspAsnLeuArgMetLysAsnProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660	QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
DB	2417	GATAATTTAAGGATGAAGATCAACCTTTTAGACTTGAATGGTATCCATCCACACCAAT	2476	DB	3497	ACCATTCTCTTCTTCCCTTCTGGAGCCATGTGTCTCTGGCCCTCTTTGTGGTACAA	3556
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680	QY	1021	IleAsnLysAlaLysSerArgProLysArgAlaGlyLysProArgValLysArgProGlnLeu	1040
DB	2477	TTCAGGGATATGACAGCATCTGAGAAGCATGGGCTACTTTGTAGAGTCTCTGGGGCC	2536	DB	3557	ATCAACAGGCCAAGAGCAGAGCGGAGGAGGAGCCAGGTGAAGCGCCGCGCAGCTC	3616
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700	QY	1041	MetGlnGlnValHisProLysThrProSerVal	1052
DB	2537	CCCTTCACGTGTTTGTATGCCAGTCAGTATGCCACTTGTCTGATGGTGACAGTGAAGG	2596	DB	3617	ATGCAGCAGGTTCAACCGCCCAAGACCCCTTCGGTG	3652
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720	RESULT 2			
DB	2597	GAGTACTTCCCTGAAGAGATCGCAAGCTCCGAGGGAGCTGGACAACGCGCTCTCGCTC	2656	HUMKIAAL			
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740	LOCUS			
DB	2657	GTCATCTTCAGTGACTGGTACCAACCTCTCTGTATGAGAAAAGTGAAGTTTATGATGAA	2716	DEFINITION			
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyValAlaAsnIleProAlaLeuAsnGlu	760	ACCESION			
DB	2717	NACACAAGCACTGGTGGATCGCGATACCGGAGGAGCTAACATCCAGCTCTGAATGAG	2776	VERSION			
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyPheThrLeu	780	KEYWORDS			
DB	2777	CTGCTCTCTGTGGAACATGGGTTTCAGCGATGGCCTGATGAAGGGGAGTTACCCCTG	2836	SOURCE			
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800	ORGANISM			
DB	2837	GCCAAACCATGATGATTTATGCTGTCAGGGTGCAGCATCCGAAGTTTCCAGAAGATGCG	2896	REFERENCE			
				AUTHORS			
				TITLE			
				JOURNAL			
				MEDLINE			
				REFERENCE			
				AUTHORS			

HUMKIAAL 4338 bp mRNA linear PRI 06-OCT-2001
 Human mRNA for KIAA0091 gene, complete cds.
 D42053
 D42053.1 GI:577308
 KIAA0091.
 Homo sapiens male myeloblast cell_line KG-1 cDNA to mRNA.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1
 Nagase, T., Miyajima, N., Tanaka, A., Suzuki, T., Seki, N., Sato, S.,
 Tabata, S., Ishikawa, K.-I., Kawabayashi, Y., Kotani, H. and Nomura, N.
 Prediction of the coding sequences of unidentified human genes.
 III. The coding sequences of 40 new genes (KIAA0081-KIAA0120)
 deduced by analysis of cDNA clones from human cell line KG-1
 DNA Res. 2 (1), 37-43 (1995)
 95308325
 2 (bases 1 to 4338)
 Ohara, O., Nagase, T., Kikuno, R. and Nomura, N.

QY	461	GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro	480
DB	1877	CAAGGCCACGGCAGCTCGCTCAGAGCCCTATCAGATCTCAACAGCTACAGCCCA	1936
QY	481	GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTyrProTyr	500
DB	1937	CAGGCAAGTTTGAGCCCGCCAGCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTAC	1996
QY	501	CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
DB	1997	TGCTCCAGCCCAATCTACTANGAGGAATCCGACAGTGTAAATGTCAACATCCCAAC	2056
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTyrGlnProTyrLeuProGln	540
DB	2057	GGCATGGGAGTCACAGAGAAATGTAGATAAGCCTGACTGGCAGCCCTATTTGCCACAG	2116
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTyrProTyrSerGly	560
DB	2117	AACGGAGACAACATTAAGTTGCTTCTCTACTCTCGGTCTTATGGCCTTGTGGCGC	2176
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTyrGluGlyIleAlaGln	580
DB	2177	TACTGGCCATCTCCATTTCTGTACCAAGAAAGCGGCTTCTTGGGAAGGCATTTGCTCAG	2236
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
DB	2237	GGCCATGTATCATCTACTGTGGCTTCCCCAGCAGACAGAGTCAAAAATGTGCAGAA	2296
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer	620
DB	2297	CAGACTTCACAGATAAGCTCCCATTAAGTGAAGTAAATCTCTACTCCCGCGGAAGC	2356
QY	621	LysArgValLeuTyrAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
DB	2357	AACAGAGTTCTCTGGGATCAGTACCACACCTCCGCTATCCACTGGCTATTTCCCGAGC	2416
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisIleHisThrAsn	660
DB	2417	GATAATTTAAGGATGAAGATGACCTTTAGACTGGAATGTCATCACATCCACACCAAT	2476
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
DB	2477	TTCAGGGATATGTACCAGCATCTTGAGAAGCATGGGCTACTTTGTAGAGTCTCTCGGGGCC	2536
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
DB	2537	CCCTTCACGTGTTTGTATGCCAGTCAGTATGGCACTTTGCTGATGTGGACATGAGGAG	2596
QY	701	GluTyrPheProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeu	720
DB	2597	GAGTACTTCCTCAAGAGATCGCAAGCTCCGGAGGACGTGGACAACGGCCTCTCGCTC	2656
QY	721	ValIlePheSerAspTyrPyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
DB	2657	GTCACTCTCAGTCAGTGGACACACTTCTGTATGAGAAAAGTGAAGTTTATGATGAA	2716
QY	741	AsnThrArgGlnTyrPyrMetProAspTyrGlyValAlaAsnIleProAlaLeuAsnGlu	760
DB	2717	AACACAAGGAGTGGTGGATGCGGATACCGGAGGAGCTAACATCCAGCTCTCAATGAG	2776
QY	761	LeuLeuSerValTyrPyrAsnMetGlyPheSerAspGlyLeuTyrGluGlyPheThrLeu	780
DB	2777	CTGCTCTCTGTGTGGAACATGGGCTCAGCGATGGCTGTATGAAGGGAGTTTCAACCTG	2836
QY	781	AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
DB	2837	GCCAAACCATTCATATGATGCTCAGGTCAGCATCGCAAGTTCACAGAAGATGCG	2896
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
DB	2897	GTCTGTATACACAGACTTTCAAGGACCAAGGATTTGAGGTTTAAAGCAGGAAACAGCA	2956
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
DB	2957	GTGTTTGAACACGTCCTCCATTTTGGGACCTTATCAGATCCAGCTGAGGGTGGAGCGCG	3016
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
DB	3017	ATTGTACTGTATGGGACTCCAATTTGTTGGATGACAGTCCAGGACAGAAAGGACTGCTTT	3076
QY	861	TyrLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSer	880
DB	3077	TGCTTCTTGATCCCTCTCCAGTACATCATGATGGGTGACACCGCTAGCTCAGT	3136
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
DB	3137	CACCTCTGGGAACCCAGCCAGCCCTCCAGTGGAGCAGGCTCAGTCACTCCAGAGAGATG	3196
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
DB	3197	GAAGGAACCATCTTCACTCGGTACTCCAAGTTCTGGAGGCCATTTGGGAGACCCAAAA	3256
QY	921	ProArgProLeuProAlaCysProArgLeuSerTyrAlaLysProGlnProLeuAsnGlu	940
DB	3257	CCTCGGCTCTACAGCTGTCCAGCTTGTCTTGGGCCAAGCCACAGCCTTTAAACGAG	3316
QY	941	ThrAlaProSerAsnLeuTyrLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal	960
DB	3317	ACGGGCCCCAGTAGTACCTTTGGAAACATCAGAAGCTACTCTCCATTGACCTGGACAAGGTG	3376
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
DB	3377	GTCTTACCACAACTTTCATCGATCGCCCTCAAGTGAAGGCCCTTGTCCTCCCTGGAGAGC	3436
QY	981	GlyAlaTyrAspTyrProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln	1000
DB	3437	GGCGCTCGGCACATCTCTGGAGGATCATGCTTGGCGCTACAAACAGAGAGTGGGCCAG	3496
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGln	1020
DB	3497	ACCATTCCTGCTCTTGTCTTCTTGGAGCCATGGTGTCTGCTGCTTCTTGTGTGATAAA	3556
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
DB	3557	ATCAACAAGGCCAAGAGCAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC	3616
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
DB	3617	ATGCACAGGTTTCAACCGCCCAAGACCCCTTCGGTG 3652	
RESULT 3			
AF078105		4198 bp mRNA linear ROD 19-NOV-1998	
LOCUS		Cricetulus griseus site-1 protease of sterol regulatory element	
DEFINITION		binding proteins mRNA, complete cds.	
ACCESSION		AF078105	
VERSION		AF078105.1 GI:3892203	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			

TYGAGVRSQVKGCCPARGSPVACANTLLSVTKQBELVNPASVQALIASA
RUPVGNPFGUCKLDLRAVQITLSSVYKQALSPSTLSTECPTWMPYQKQF115
GMPTLVNVTLLMGCTGCTIVKAPRPEPLGNGTQVQVQVQVQVQVQVQV
SVTKKASWEEGIAQIIMTVASAPKELKNGRUTSVLP121LTPTPSRQV1
WQOYNLAVPGVTPDNILKQNPDLQWDVHNPEDMAVHLRSKVFYEVGACPF
TCDATQVGLTLMDSEERYPEETAKLREDVQNGLSLVFSDYNTSVMAKVKFE2
NTQWQMPDQGANVALNELLVNMWSSDGLYEGEELANHDYVAGSGC2ARPE
DGVVLTQVQKQGLVEKQETAVNVPVILGQVQIAPBGGRTVLVYGDNSGLDHSRPE
KDCVFLDALQVLTGVTVPPLSLSHGNRQKPPSQGLAPPERMEGNHLRVSKLEA
HLGDKPRLPACPLSHWAKPOKPLNENAPSNLWKHKQLLSLDLQVLPNPRNRQV
RPLSPGSGEANDIPGIMPGRYNEVQGTIPVFAFLGMALVAFVQVQISAKSRPRK
RRPRKBPOLAOOAHBAPTSPV*

Db 2218 CATACCTCCAGCAGTGGCCATCAAGGTGAAGATCATTCCTCCACCCTCTCCGAGC 2277
QY 621 LysArgValLeuTyrAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg 640
Db 2278 AGAGAGTCTCTGGGACCACTACCAACACTCCGCTACCCACCCGGCTACTTCCCAGG 2337
QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisLeuHisThrAsn 660
Db 2338 GACAACCTTCGGATGAAGATGATCTTTAGACTGGAATGGGACCACTCCACACCAAC 2397
QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
Db 2398 TTCAGGAGATGATACAGCATCTCGCAGCATGGCTACTTTGTGAGGTGCTTGTGTC 2457
QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
Db 2458 CCATTCACATGCTTTCAGCCAGCCAGTACGGCAGCTCTGTTATGGTGACAGTAGGAA 2517
QY 701 GluTyrPheProGluGluLeuAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
Db 2518 GAGTACTTCCCTGAGGAGATGCTTAAGCTGAGGAGGAGCTGGACATGGCCTTCCCTT 2577
QY 721 ValIlePheSerAspTyrTrpAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
Db 2578 GTCGTCTTCAGTGACTGGTACACACTTCTGTATGAGAAAGTGAAGTTTACGATGA 2637
QY 741 AsnThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlu 760
Db 2638 ACACAAAGCAGTGGTGGATGCGAGATACATGAGGAGCCCAAGCTCCAGCTCTAAACGAG 2697
QY 761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
Db 2698 CTGCTCTGTGTGGAACATGGGTTTCAGTGAGCGGCTGTATGAAGGGAGTTGGCCGT 2757
QY 781 AlaAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
Db 2758 GCAAAACAGCAGATGATGATGATGCGGGTGCAGCATGCCAGTTTCCAGAAGATGT 2817
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
Db 2818 GTGGTATCACAGACACTTTCAGGACCAAGGATGGAAGTCTTAAACAAGACAGACGA 2877
QY 821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
Db 2878 GTTGTGCAATGTCCTCCATCTCGGGCTATATCAGATCCAGCTGAAGTGGAGCCGG 2937
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
Db 2938 ATTGTGCTGTATGGAGACTCCAACTGCTTGGATGACAGTCACAGACAGAAGGACTGCTT 2997
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2998 TGCCTTCTGGATGACATCTCTCAGTACACATCTATGGTGTGACCCCTCCAGCCCTCAG 3057
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3058 CATTCAGGAACCGGACGCCACCCAGCGGGCTGGCTGGCCCTCTCTGAAGGATG 3117
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3118 GAAGGAACACCTTCTCATGCTACTCCAAGTCTTGGAGCCCACTTGGAGACCCGAAA 3177
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3178 CTCGGCCCTTCCAGCCCTGTCACACTTGTGTGGGCCAAGCCACAGCTTTGAATGAG 3237
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
Db 3238 ACGGCACCCAGTAATCTTTGGAAACACCAAGAGCTCTCTCCATTGACCTGGACAAAGTA 3297
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980

Db 3298 GTGTTACCAACCTTTCGCTCAAAATCGCCCTCAAGTGAAGACCTTTGCTCCCTGGAGAAGT 3357
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
Db 3358 GGTGCTTGGGACATCTCTGGAGGATCATGCTTGGCGCTTCAACACGAGGATGAGCCAG 3417
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln 1020
Db 3418 ACCATCCCTGTTTTCCTTCCCTTGGAGCATGCTGGCCCTGCTTCTTCGTGGTACAG 3477
QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
Db 3478 ATCATGAAGCCCAAGACCGCGCGAGGAGGAGCCAGGCAAGCGTCCACACTT 3537
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3538 GCACAGAGCCCAACCTGCAAGGACCCCGCTCAGT 3573

RESULT 5
AF094821 3895 bp mRNA linear ROD 25-APR-1999
LOCUS Rattus norvegicus subtilisin/kexin isozyme SKI-1 precursor, mRNA,
DEFINITION complete cds.
ACCESSION AF094821
VERSION AF094821.1 GI:4679094
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 3895)
AUTHORS Seidah,N.G., Mowla,S.J., Hamelin,J., Mamarbachi,A.M., Benjannet,S.,
Tours,B.B., Basak,A., Munzer,J.S., Marcinkiewicz,J., Zhong,M.,
Barate,J.C., Iazure,C., Murphy,R.A., Chretien,M. and
Marcinkiewicz,M.
Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed
proprotein convertase with a unique cleavage specificity and
cellular localization
Proc. Natl. Acad. Sci. U.S.A. 96 (4), 1321-1326 (1999)
JOURNAL 99145548
MEDLINE 9990022
REFERENCE 2 (bases 1 to 3895)
AUTHORS Seidah,N.G., Mamarbachi,A.M., Hamelin,J. and Chretien,M.
Direct Submission
TITLE Submitted (24-SEP-1998) Biochemical Neuroendocrinology, Clinical
JOURNAL Research Institute of Montreal, 110 Pine Ave. West, Montreal, QC
H2W 1R7, Canada
FEATURES
source
1. .3895
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/cell_line="PC12"
/tissue_type="adrenal glands"
418. .3576
/note="subtilisin-like serine proteinase; cleaves after
hydroxylated amino acids such as Thr and possibly Ser"
/product="subtilisin/kexin isozyme SKI-1 precursor"
/protein_id="AAD27011.1"
/db_xref="GI:4679095"
/translation="MKLVNLIWLLLVLLCGKKHGLDRLGKKAFAKAPCSCSHLTK
VEESFTVVEYIVAEVGYTAKARNSFISSALKSSEVDNWRILIPNNSSDYPDFE
VQIKKQKAGLLTLEDHPIKRVTPQRVRSIKFAESDPIVPCNETRQKWSR
PLKRASLSSGFWHATGRHSRRLLRPRQVATQADVLVMQMGYTGANVRVFD
TGLSEKHPFKNVKERTNWTNERTLDGLGHTFGVAVTASMECCGCFAPDAELHFR
VTNNQVSYTNFLDAFNAILKMDVLNLSIGGPDFMDHPFDVKVWELTANNVMS
ATNGDPLGTLNPNADQMDVIGVGIGDFEDNIARFSRRGTTWELPGLVGRVKPDIV
TVGARGSGVGGCCALSGTSVSPVAGATVLLVSTVQKRELNPASVKQALLASA
RRLPGVMEFQGHKLURLRAYQILSYAPQASLSPSIDLECPFWMPFCSPITYG
GPTIYNVTILMGVTRIVDKPEWRYPQNDNIYAFVSFYSVLPWPMWGLAIISI
SVTKKAASWEGIAQGHMITVASPAETELKNGAETHTVKLPIKVKIITPTPRSKRLV

WDQYHNLRYPPGYPFRONLRKMKNDPLDWNQDHHVHTNFRDMYQHLRSMGYFVEVLGAPF
 TCDAQYQTLGLMVDSEEEYFPEETIAKLRDNDVGLSLVVFSDWNTYSVMKRVFYDE
 NTQWMPDPTGLMANYPALNELLSSWNMGFSGLYEGETFALANDMYIYASGCSIAKPEE
 DGVITQTEKDGLEVLKQETAVNDVPTILGIYIPAEQGGRIYLYGDSNCLDSDHRO
 KDCFWLLDALQYTSYGVTPPSSHSNGNRQRPSPGAGLAPPERMEGNHLHRYISVLEA
 HLGDPKPRPLPACPHLSWAKPQPLNAPSNLWKHOKLLSIDLKVLPNFRSNRPQV
 RPLSPGESGAWDIPGIMPGRYNOEVGOTIPVFAFLGANVALAFFVVOISRAKSRPKR
 RRPRAKRPOLAOQARHAPTPSV"

sig_peptide 418..468
 mat_peptide 469..3573
 misc_feature /product="subtilisin/kexin isozyme SKI-1"
 1069..1071
 /note="active site"
 1123..1125
 /note="glycosylation site"
 1162..1164
 /note="active site"
 1330..1332
 /note="glycosylation site"
 1429..1431
 /note="active site"
 1657..1659
 /note="active site"
 1957..1959
 /note="glycosylation site"
 2599..2601
 /note="glycosylation site"
 3232..3234
 /note="glycosylation site"
 3421..3480
 /note="transmembrane-region site"
 3481..3573
 /note="cytosolic tail; unclassified site"

BASE COUNT 952 a 991 c 1055 g 897 t

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 3895
 Score: 5476.00 Matches: 1017
 Percent Similarity: 98.57% Conservative: 20
 Best Local Similarity: 96.67% Mismatches: 15
 Query Match: 97.49% Indels: 0
 DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x AF094821 (1-3895)

QY 1 MetLysLeuValAsnIleThrLeuLeuValValLeuLeuCysGlyLysLysHis 20
 DB 418 ATGAAGCTCGTCAACATCTGGCTTCTTCTGCTGGTTCCTGCTGGGAAAGCAT 477
 QY 21 LeuGlyAspArgLeuGluLysLysSerPheGluLysAlaProCysProGlyCysSerHis 40
 DB 478 CTGGGTGACAGCTGGGGAGAAAGCTTTTGAAGAGCCCAATGCCAGCTGTCCAC 537
 QY 41 LeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAlaPhe 60
 DB 538 CTGACTTTGAAGTGGAAATCTCCTCAACTGTGGTGGAAATATGAATATATGTGGCTTTC 597
 QY 61 AsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSerSer 80
 DB 598 AACGGATATTCACAGCCAAAGCTAGAACTCATTTATTTCAAGTCTCAAAAAGCAT 657
 QY 81 GluValAspAsnTrpArgIleIleProArgAsnProSerSerAspTyrProSerAsp 100
 DB 658 GAAGTGGACACTGGAGANTATACCTCGGAACAACCCATCTAGTACTACCTAGTAT 717
 QY 101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuThrLeuGluAspHis 120
 DB 718 TTTGAGGTGATTCAGATMAAGAGAGCAGGAGCGGGCTGCTCAGACTTGAAGATCAC 777
 QY 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 DB 778 CCAACATCAAGCGGTGACACCCAGCGGAAAGTCTTTTGGTTCCTCGAAGTTTGTGAA 837

QY 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSerArg 160
 DB 838 TCCGACCCCATTTGTCCTCTGTAATGAGCCCGGTGGAGCCAGAGTGGCATCAACGT 897
 QY 161 ProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheTrpPheIleAlaThrGlyArgHis 180
 DB 898 CCCCCTGAAAAGAGCCAGTCTCTCCCTGGGCTCTGGATTTCTGGCATGCAACAGGAGCAT 957
 QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 DB 958 TCAAGTCGACGATTTGCTGAGAGCCATTCCTCGCCAGGTTGCCACACATTCGACGCAGAT 1017
 QY 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValAlaPheAspThrGly 220
 DB 1018 GTGCTTTGGCAGATGGGATACACAGGTGCTAATCTCAGGGTTGCCGCTTTTCATACATGGG 1077
 QY 221 LeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 DB 1078 CTCAGTGAGAGCATCCACATTTCAAGAATCTGAAGGAAGAACCACTGACCAATGAG 1137
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 DB 1138 CGGACCTGGACCATGGGCTGGCCATGCCACATTCGTCAGGTGTGATTGCCAGCATG 1197
 QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 DB 1198 AGAGAGTCCCAAGGATTTGCCCATGCAAGCTGCACATCTTCAGGGTCTTTACCAAC 1257
 QY 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
 DB 1258 AATCAGGTGCTTACAGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1317
 QY 301 IleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHisProPheValAsp 320
 DB 1318 ATGGAGCTTCTGAACCTTACATCGGTGGGCTGACTCATGATGATCACCCTTTGTGAC 1377
 QY 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 DB 1378 AAGGTATGGGAATTAACAGCGAACAAATTAATCATATGTTCTCTATTGGCAATGATGGA 1437
 QY 341 ProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyGly 360
 DB 1438 CCTCTCTATGACATCTGTAATACCCCTGCTGATCAGATGATGATGATGATGATGATGATG 1497
 QY 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeu 380
 DB 1498 ATTGACTTTGAAGACAAACATCGCCGCTTCTCTCCAGGGGAATGACTACCTGGGAATA 1557
 QY 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGly 400
 DB 1558 CCGGAGGCTATGCTCTGTGAAGCTGACATTCACCTATGCTGCTGAGTGGGGT 1617
 QY 401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 DB 1618 TCTGGTGTGAAGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1677
 QY 421 AlaGlyAlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
 DB 1678 GCTGGGGCTGTCACCTTTGTAGTAAAGCAGTACAGAGCGGGAGTAGTGAATCCTGCC 1737
 QY 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGlu 460
 DB 1738 AGTGTGAAGCAAGCTTTGTAGTATGATCAGCCCGGAGACTTCTGCTGCTCAACATGTTTGA 1797
 QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480
 DB 1798 CAAGGCCATGCCAAGTTGGATCTACTTGGAGCCTATCAGATCTCTCAGCAGCTATAACCG 1857
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
 DB 1858 CAGGCGAGCCTGATCTCTACATACCTGACCTGACTGAGTGTCCCTACATGTGGCCCTAC 1917

BASE COUNT 915 967 891 t

AIGNCPYXGTLNPADQMDWTLGVGGIDFEDNIRAFSSRGMTTWELPGYGRVXPEDVL
 TYGCRSGAGGCLSGTSSVYAGANVTLVTKRELVNFWPSPVSKQALIASA
 PRQVNFNPFQDQGLKQVYKQVLSKTKQASLSPSIDTECPMTWSPGQITGIG
 GMTYVNTVTLQUCGQVQVYKQVLSKTKQASLSPSIDTECPMTWSPGQITGIG
 SVYKASMEGIIAGQIHMTVASNAPETLILQKQNDIRAFVSSVYMFWSGLKSLV
 WQYKHNLYVSPQPNDRIMKNDPLDQNGDVNHPDQVTLNKGATVYVILGAE
 TCDATQVGTLLVLDSEERYEPETAKLREDVNGLSLVAFSTNDYVYVILGAE
 NTBOMQPDGTALNPALENLVNLWNGSPGLYSGRELVFANNDMYVYASGCSIAKPE
 KDGVTLQKQGVLYKQETAVNPNVILGLVYLPSEGGGRVLYXGSDNSVSKR
 KDCVFLDLAQVLYSGVTPVPSLSHSGNRORPPSCAGLAPERMEGNHLHRYSLVLA
 HLGDPKPSLPACPHVSKMAKPOPLNEPNTSLWKHOKLLSIDDKVYLPNPSNRQV
 RPLSPGSEANDIPGGIMPRYNEVGQITPVPFAFLGAMVALAFVVOILSKAKSRKR
 RRPRAKPOLQAQAHPATPSY"

QY 601 GlnThrSerThrValLysLeuProLysValLysValLysLeuProThrProArgSer 620
 DB 1808 CACACATCCACCGTGAAGTGCCTCAAGGTGAAGATCATTCACCCCTCCCTCGGAGC 1867
 QY 621 LysArgValLeuThrPheGlnHisAsnLeuArgTyrProGlyTyrPheProArg 640
 DB 1868 AAGAGAGTCTCTGGGACCAAGTACCACAACTCCGCTACCCACTGGCTACTTCCCGAGG 1927
 QY 641 AspAsnLeuArgMetLysAsnAspProLeuAspTyrAsnGlyAspHisLeuHisThrAsn 660
 DB 1928 GACAACTTGGGATGAAGATGACCTTTAGACTGGATGGCAGCCACACGTCACACCAAC 1987
 QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
 DB 1988 TTCAGGAGATGATACAGCATCTCGCGAGCATGGCTACTTCTGGAGGTGCTCGCGGCC 2047
 QY 681 ProPheThrCysPheAspAlaSerGlnTyrGlnTyrLeuLeuMetValAspSerGluGlu 700
 DB 2048 CCAATTCACATGTTTGGCCACACAGATGGCACCTTTCGCTGGTGGACATGAGGAA 2107
 QY 701 GluTyrPheProGluGluLeuLysLeuArgAspValAspAsnGlyLeuSerLeu 720
 DB 2108 GAGTACTTCCCTGAGGAGATGCTAAGCTGAGGAGGATGTGCACATGSCCTTCCCTC 2167
 QY 721 ValIlePheSerAspTyrPheAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
 DB 2168 GTCATCTTCAGTGACGTGATACAACTCTCTGTTATGAGAAAGTGAAGTTTATGATGA 2227
 QY 741 AsnThrArgGlnTyrPheMetProAspThrGlyAlaAsnLeuProAlaLeuAsnGlu 760
 DB 2228 AACACAGGAGTGGTGGATGCGCAGACACCGGAGGAGGATGACATCCAGCTCGAATGAG 2287
 QY 761 LeuLeuSerValThrPheAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
 DB 2288 CTGCTGCTGTGTGGAACTGGGTTCAGTGACGCGCTATATCAAGGGAGTTTGTCTG 2347
 QY 781 AlaAsnHisAspMetTyrThrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
 DB 2348 GCAACCATGACATGATGATGCTGCGGTGCGGATGAGTGTGAGGTCTTAACAAAGAGACGA 2467
 QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
 DB 2408 GTGCTGATCACAGACTTTCAGGACCAAGGATGAGGTCTTAACAAAGAGACGA 2467
 QY 821 ValValGluAsnValProLysLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840
 DB 2468 GTTGTGGAAATGTTCCCATTTTGGGCTTATCAGATTCATCCTCAAGGTGAGCGCG 2527
 QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
 DB 2528 ATCGTGTGTATGGAGACTTCAACTGCTTGGATGAGATGACAGTACAGAGAGGACTGCTT 2587
 QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
 DB 2588 TGGCTTCTGGATCGCTCTTCAGTACACATCTATGGGTGACCCCTCCACCCCTCAGC 2647
 QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
 DB 2648 CATTGAGGAACCGCGAGCGCCACCTAGCGGAGCGGCTTGGCCCTCTCTGAAAGGATG 2707
 QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
 DB 2708 GAAGGAACACCTCTCATCGGTACTCCAAAGTTCTTGAAGCCCACTTGGAGAGACCGGAA 2767
 QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
 DB 2768 CCTCGGCCCTGCGACCTGCTCCACATTTGTATGGGCCCAAGCCACAGCCCTTTGAATGAG 2827
 QY 941 ThrAlaProSerAsnLeuThrLysHisGlnLysLeuLeuSerIleAspLeuAspLysVal 960
 DB 2828 ACGGCCACCCAGTAATCTTTGGAAACATCAGAAGCTGCTCTCCATTGACCTGGACAAAGTA 2887
 QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980

DB 2888 GTGTACCCCAACTTCGATCCAATCGCCCTCAAGTGAGACCTTTGTCCCTCGGAGAGT 2947
 QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
 DB 2948 GGTGCTGGGACATCTCTGGAGGATCATCCCTGGCGCTACCAACGAGGAGTGGGACAG 3007
 QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln 1020
 DB 3008 ACCATCCCGCTCTCGCCCTTCCTCGGAGCATGGTGGCCCTGGCTCTTTTGGTACAG 3067
 QY 1021 IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu 1040
 DB 3068 ATCAGCAAGGCCAAGACCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3127
 QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
 DB 3128 GCACAGAGGCCCAACCTCGAAGGAGCCCATCATGAGT 3163
 RESULT 7
 AF094820 3788 bp mRNA linear ROD 25-APR-1999
 AF094820
 LOCUS Mus musculus subtilisin/kexin isozyme SKI-1 precursor, mRNA, complete cds.
 DEFINITION AF094820.1 GI:4679092
 ACCESSION AF094820
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus.
 Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 3788)
 AUTHORS Seidah,N.G., Mowla,S.J., Hamelin,J., Mamarbach,A.M., Benjannet,S., Toure,B.B., Basak,A., Munzer,J.S., Marcinkiewicz,J., Zhong,M., Barale,J.C., Lazure,C., Murphy,R.A., Chretien,M. and Marcinkiewicz,M.
 TITLE Mammalian subtilisin/kexin isozyme SKI-1: A widely expressed proprotein convertase with a unique cleavage specificity and cellular localization
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (4), 1321-1326 (1999)
 MEDLINE 99145548
 PUBMED 9990022
 REFERENCE 2 (bases 1 to 3788)
 AUTHORS Seidah,N.G., Mamarbach,A.M., Hamelin,J. and Chretien,M.
 TITLE Direct Submission
 JOURNAL Submitted (24-SEP-1998) Biochemical Neuroendocrinology, Clinical Research Institute of Montreal, 110 Pine Ave. West, Montreal, QC H2W 1R7, Canada
 FEATURES
 source 1..3788
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /cell_line="AtT20"
 8..3166
 /note="subtilisin-like serine proteinase; cleaves after hydroxylated amino acids such as Thr and possibly Ser"
 /codon_start=1
 /product="subtilisin/kexin isozyme SKI-1 precursor"
 /protein_id="AAD27010.1"
 /db_xref="GI:4679093"
 /translation="MKLVSTWLLVLLVLCGRKHLGRLGTRALEKAPCPSSHLTLK VFSSFTVEYIVAFNGYFTAKRNSFISSALKSEVENWRIIPRNSSDPDFE VTIQIKERKAGLLTLEDHPNIRKVTQPKRFSLKFAESNPVPCNWTSPQKWSR PLKRAISLGSFGFWHATGRSSRLRLRAIPROVAOTLQADVLWOMGYTGANVAVFD TGLSEKHPKKNVKRTNNTERTLDGLGHGTFFVAGVTASMECCOGFADAEHLFR VFTNQVSYTSWFLDAFNALKKMDYLNLSLGGPDMHDPFVKVWELTANNVIMVS AICNGDPLGLTNLPADQMDVIGVGDFEDNIARFSRGMWTLWELPGGVKRPDI VTCGAGRGSGVGGCCALSGTSVAPVAVAGATLVLLVTKRELVPNPASVKQALIASA RPLPGVMEQGHGKLDLLRAYQILSSYKPSQSLSPYIDLTECPYMWPCSQPIYVG GMPTIVNVTILNGMGTGRIVDKPEWRPYPONGDNIEVAFVSVPWSPGSLAISI SVTKKAASNEGIAQGHIMITVASPAETELHSGAEHTSTVKLPKVIITPPRSKRL WQYHNLRTPPGTPRDNLRMKNDPLDNGDHHVHTNFRDMYQHLRSNGTFFVEIGAPF TCFDATQYGTLLLVDSSEBYFPETIAKLRRRDVNGLSLVISDWNTYSMRKVKFYDE

QY	501	CysSerGlnProIleTyrGlyMetProThrValValAsnValThrIleLeuAsn	520
Db	1508	TGCTCCACGCTATCTACTATGAGGAATGCCAACAATGTTAATGTCAACCTCCTCAAT	1567
QY	521	GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln	540
Db	1568	GGCATGGGGCTCACAGGAAGAAATGTGGATAAGCCTGAGTGGCGACCCCTATTACACAG	1627
QY	541	AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly	560
Db	1628	AATGGAGACAATGAAGTGCCCTCTCTACTCTCAGTGTGGGCCCTGGTCAGGT	1687
QY	561	TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln	580
Db	1688	TACCTTGCCATCTCCATTTCTGTGACCAAGAAGGAGCCTCTCTGGGAAGGCATCGCTCAG	1747
QY	581	GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu	600
Db	1748	GGCCCATCATCATGACAGTGGCGTCCCCAGAGACAGAGATTACACAGTGTGGCGAG	1807
QY	601	GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer	620
Db	1808	CACACTTCACCGTGNAGCTGCCATCAAGGTGAAGATCTTCCACCCCTCTCGGAGC	1867
QY	621	LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg	640
Db	1868	AAGAGAGTCTCTGGGACCACTACCAACCTCCGCTACCCACCTGGCTACTTCCCGAGG	1927
QY	641	AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn	660
Db	1928	GACAACTTCGGATGAAGAAATGACCTTTAGACTGGAAATGGGACCACTCCACACCAAC	1987
QY	661	PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla	680
Db	1988	TTACGGGACATGTACCAGCATCTGCCACCATGGCGTACTTCTGTGGAGGTGCTGGCGCC	2047
QY	681	ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu	700
Db	2048	CCATTTCACATGTTTGGACCCACACAGTAGTGACCTTCTGCTGTGTGACAGTAGGAA	2107
QY	701	GluTyrPheProGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu	720
Db	2108	GAGTACTTCCCTGGAGAGATTCTAAGCTGAGGAGGGATGTGGACAATGGCCTTCCCTC	2167
QY	721	ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu	740
Db	2168	GTCATCTTCAGTGTGTACAACTTCTGTATGAGAAAGTGAAGTTTATGATGAA	2227
QY	741	AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu	760
Db	2228	AACACGAGCAGTGTGGATGCCAGACACCGGAGGAGCAACATCCACGCTCTGAATGAG	2287
QY	761	LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu	780
Db	2288	CTGCTGTCTGTGGAACATGGGGTTCAGTCAGCGCCTATATGAAGGGAGTTTGCCTG	2347
QY	781	AlaAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGly	800
Db	2348	GCAAACCATGATGATGATGCTGGTGGGGTGCAGCATGCCAAGTTTCCAGAAGATGCC	2407
QY	801	ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla	820
Db	2408	GTCTGTATCACACAGCTTTCAGGACCAAGGATGGAGGTCTTAAACAAGACAGACGA	2467
QY	821	ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg	840
Db	2468	GTGTGGAAATGTTCCCATTTTGGGGCTTTATCAGATTCATCTGAAGTGGAGGCCGG	2527
QY	841	IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe	860
Db	2528	ATCGTGCTGTATGGAGACTCCAACTGCTTGGATGACAGTACAGACAGAAAGACTGCTTT	2587
QY	861	TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer	880
Db	2588	TGGCTTCTGGATGCGCTCTTCAGTACACATCTATGGCTGACCCCTCCAGCCTCAGC	2647
QY	881	HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet	900
Db	2648	CATTGAGGAAGCGGAGCGCCACCTAGCGGAGCGGCTTGGCCCTCTCTGAAGAGATG	2707
QY	901	GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys	920
Db	2708	GAAGGAACACCTCCATCGGTACTCCAAAGTTCTTGAAGCCACCTTGGAGACCCGAAA	2767
QY	921	ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu	940
Db	2768	CCTCGGCCCTCGCAGCCTGTCCACATTTGTATGGGCCCAAGCACAGCCTTTGAATGAG	2827
QY	941	ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLysVal	960
Db	2828	ACGGCACCCAGTAATCTTTGGAACATCAAGAGCTCTCTCCATTGACCTGGACAAGTA	2887
QY	961	ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer	980
Db	2888	GTGTATACCACTTTCATCCCAATCCCTCAAGTGAGAGACCTTTGCTCCCTGGAGAGT	2947
QY	981	GlyAlaTrpAspIleProGlyIleMetProGlyArgTyrAsnGlnValGlyGln	1000
Db	2948	GGTGCCTGGACATCTCTGGAGGATCATGCTGGCGCTACAAACAGGAGTGGGACAG	3007
QY	1001	ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPheValValGln	1020
Db	3008	ACCATCCCGTCTTCCGCTTCTCGGAGCATGTTGGCCCTGGCCCTTCTTGTGGTACAG	3067
QY	1021	IleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGlnLeu	1040
Db	3068	ATCAGCAAGCCCAAGACCGCGCGAAGCGGAGGAGGCCCAAGCGTCCACAACTT	3127
QY	1041	MetGlnGlnValHisProProLysThrProSerVal	1052
Db	3128	GCACAGCAGCGCCACCCCTGCAAGGACCCATCATGTG	3163
RESULT 8			
BC011533			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REMARK			
COMMENT			

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
 Series: IRAC Plate: 5 Row: e Column: 7.
 Location/Qualifiers

FEATURES

source.

1. 4204
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="MGC:6457 IMAGE:2615736"
 /tissue_type="Mammary tumor, C3(1)-tag model. Infiltrating ductal carcinoma, 5 month old virgin mouse."
 /clone_lib="NCI CGAP_Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-Sport6"
 406.3561-1
 /codon_start=1
 /product="Similar to membrane-bound transcription factor protease, site 1."
 /protein_id="AAH11533.1"
 /db_xref="GI:15079377"
 /translation="MKLVSTWLLVVLVLLGKRHLGDRIGLRALEKAPCPSCSHLLIK VIGKEKQAGLLTLEDPNKKRTPQKRVFSLKSKSEVENRIIPKNNPSSDIQSDRE PLKRSLSIGSGFWHATGRHSSRLRLRAIPROVAQTLQADVLWMCYTCANRVAYED TGLSEKHPFKNVKERTNWTNERTLDGHHGTGTVAGVVIASMECCGAPDAELHIFR VFTNNQVSTWFLDAFNALIKKMDVNLNLSIGGPDFMDHPFVFWELTANNVIVS AIGNDGLYXTLNNDPADQMDVIGGVDIFEDNIAFSSRGMTTWELPGGYKRPDIIV RYLGAGVSGVGGCRALSGTSVAPVAGAVTLLVSTQKRELNPASVQKQALIASA RRLPGVMFEQHGKLDLRAIQILSSYKQASLSPSYIDITECPYMPYCSQPIYGG MPTIVNVTLINGMGVTRIVDKPEWRPYLPQNGDIEVAFSYSVLMPWSYLAISIS VTKKASWEGIAQGHIMITVASPAETELHSGAHTSTVKLPKIKLIPTPPRSRLVM CDYHNLRYPPGYPRDLNRMKNDPLDNGHDVHTNFRDMYQHLRSNGYFEVLGAPFT OFATQVGTLLLVDEEYEPPEETAKLRDNDGLSLVIFSDWYNTSVMRKVYEDEN TROMWPDGGANIPALNELLISVNNMGFSDLGYECEFVLANDHMYVASGCSAKFPED GVITOTFKDOGLEVLKQETAVVNPVILGYOIPSEGGGRIVLYGDSNCLDSDSHROK DFWLLDALQYTSYGVTPPSLSHNSGNRORPPSGAGLAPPERMEGNHLHRSKYLEAH LGDPKRLPACPHLSNAKPOPLNETAPSNLWKOKLLSIDLQVVLNFRSNRPQVR PLSPGSGAGIPGGIMPEGRNTEVQGTIPVFAFLGAWVALAFVQVQISKAKSRPKR RPRAKRPLQAQAQAPARTPSV"

CDS

BASE COUNT 1029 a 1061 c 1136 g 978 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 4204
 Score: 5412.50 Matches: 1007
 Percent Similarity: 98.10% Conservative: 25
 Best Local Similarity: 95.72% Mismatches: 19
 Query Match: 96.36% Indels: 1
 DB: 10 Gaps: 1
 US-09-830-837-6 (1-1052) x BC011533 (1-4204)

101 PheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAspHis 120
 706 TTTGAGTGATTCAGATAAAGAGAGACAGACGGGGCTCTCACACTTGAAGATCAC 765
 121 ProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAlaGlu 140
 766 CCCAACATCAAGCGGTGACACCCCGAGCGAAAGTCTTCGPTCCCTCAAGTTTGGTGA 825
 141 SerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTyrPheSerArg 160
 826 TCCAAACCCATCGTCCCTTAATGAACCCCGTGGAGCAGAGTGGCAGTATCATCGT 885
 161 ProLeuArgArgAlaSerLeuSerLysGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 886 CCCCTGAAAGAGCCAGTCTCTCCCTGGGCTCTGGATCTTGGCATGCAACAGGACAT 945
 181 SerSerArgArgLeuLeuAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 946 TCAAGTCGCGATTGCTGAGAGCCATCTCCGCGAGTCCGCGACAGACTCGAGCGAGAT 1005
 201 ValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
 1006 GTGCTGTGGCAGATGGGATACACAGGTGCTAATGTCCAGAGTTGCTGTTTTGATCTGG 1065
 221 LeuSerGlnLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsnGlu 240
 1066 CTCAGTCAGAGACATCCGCATTTAAGAAATGTGAAGAGAGAACCACTGGACCAATGAG 1125
 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 1126 CGGACCTTGGATGATGGGCTAGGCATGGCAGTCTGTCAGGTGTGATTTGCCAGCATG 1185
 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 1186 AGGAGATGCCAAGGATTTCTCCAGATGACAGTGCACATCTTCAGGCTTTTACCAC 1245
 281 AsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 300
 1246 AATCAGTGTCTTACACATCTTGGTTCCTGGATGCCCTTCAACTATGCTATTTGGTGA 1305
 301 IleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheValasp 320
 1306 ATGGACGTTCTCAACCTTACATCGTGGCGGACATCATGGATCATCCGTTTGTGTGAC 1365
 321 LysValTrpGluLeuThrAlaAsnValIleMetValSerAlaIleGlyAsnAspGly 340
 1366 AAGGTGTGGGAATTAACAGCTAACCAATGTAATATGTTTCTGCTATTGGCAATGATGA 1425
 341 ProLeuTyrGlyThrLeuAsnProAlaAspGlnMetAspValIleGlyValGlyGly 360
 1426 CCTCTATGCGACTCTGTAATACCTCTCATCAGATGGATGATGATGGAGTGGTGGC 1485
 361 IleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGluLeu 380
 1486 ATGACTTTGAGATTAACATCGCTGCTTTCTCCAGGGGAATGACTACCTGGGAATTA 1545
 381 ProGlyGlyTyrGlyArgMetLysProAspIleValThrGlyAlaGlyValArgGly 400
 1546 CCAGAGGCTATGCTGTGTGAAGCTGCACTGATGATGCTACCTATGCTGCTGGAGTGGGGT 1605
 401 SerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVal 420
 1606 TCCGCTGTGAAGGGGCTGCCGTGACTCTCAGGGACAGTGTGCTTCCCGAGTGTGC 1665
 421 AlaGlyAlaValThrLeuValSerThrValGlnLysArgGluLeuValAsnProAla 440
 1666 GCTGGGGCGCTACCTTTGTAGTAAAGCAGTACAGAGCGGAGCTGGTGAATCCCGCC 1725
 441 SerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValAsnMetPheGlu 460
 1726 AGTGTGAAGCAAGCTTTGATAGCTCAGCGCGGAGACTTCTCTGGGTCAACATGTCGAG 1785
 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysPro 480

1786 CAAGGTCATGGCAAGTTGGAT-----CTGGCAGCTTATCAGATCCTCAGCAGCTATAAACCG 1842
481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
1843 CAGCGAAGCCTGAGTCTTACCTAGTACATCGACCTGAGTGTCCCTTACATGTGGCCCTAC 1902
501 CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
1903 TGTCTCCAGCCTATCTACTATGAGGAATGCCAACATCGTTAATGTCAACCATCCTCAAT 1962
521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
1963 GGATGGGGCTCAGAGGAAGATGTGGATAAGCCTGAGTGGCGGACCCCTATTATACCACAG 2022
541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTyrSerGly 560
2023 AATGGAGACAACATTCAAGTGGCTTCTCTACTCTCAGTGTGTGGCCCTGGTCAAGT 2082
561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluIleAlaGln 580
2083 TACCTTGGCATCTCCATTCTGTGACCAAGAGGCAGCTTCTCTGGGAAGGCATCGCTCAG 2142
581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
2143 GGCCCATCATGATCAGTGGCGTCCCGCAGCAGACAGAGTTACACAGTGGTGGCGAG 2202
601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProProArgSer 620
2203 CACACTTCCACCGTGAAGTGGCCCATCAGGTGAAGATCATTTCCACCCCTCTCTGGAGC 2262
621 LysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArg 640
2263 AAGAGAGTCTCTGGGACCCAGTACCACAACCTCCGCTACCCACCTGGCTACTTCCCGAG 2322
641 AspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsn 660
2323 GACAACCTCGGGATGAAGATGACCTTTAGACTGGGAATGGGACCACTCCACACCAAC 2382
661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
2383 TTCAGGACATGTACCGAGTCTGGCAGCATGGGCTACTTCTGGAGGTGCTCGGCGCC 2442
681 ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGlu 700
2443 CCATTTCACATGTTTTCACGCCACACAGTATGGCACTTTCGCTGCTGGTGACAGTGGAGAA 2502
701 GluTyrPheProGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeu 720
2503 GAGTACTTCCCTGAGGAGATGCTAAGCTGAGGAGGATGTGGACAATGGCCCTTTCCCTC 2562
721 ValIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGlu 740
2563 GTCATCTTCAGTGACTGGTACACACTTCTGTATGAGAAAGTGAAGTTTATGATGAA 2622
741 AsnThrArgGlnTrpTrpMetProAspThrGlyAlaAsnIleProAlaLeuAsnGlu 760
2623 AACACAGGACAGTGGTGTGTCAGACACCGGAGGAGCAACATCCAGCTCTGAATGAG 2682
761 LeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeu 780
2683 CTGCTGTCTGTGTGAACATGGGTTCAGTGGCCCTATATGAAGGGGAGTTTGTCTGT 2742
781 AlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGly 800
2743 GCAACCATGACATGTACTATGCTGGGGTGCAGATCGCCCAAGTTTCCAGAAATGCG 2802
801 ValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAla 820
2803 GTCGTGATCACACAGACTTTCAGGACCAAGGATTGGAGGTCTTAAACACAGACAGCA 2862
821 ValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArg 840

Db 2863 GTTGTGGAATAATGTTCCCATTTTGGGGCTTTATCAGATTCCATCTGAAGTGGAGCCGG 2922
QY 841 IleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGlnLysAspCysPhe 860
Db 2923 ATCGTCTGTATGGAGACTCCAACCTCTTGGATGACAGTACACAGACAGAGGACTGCTTT 2982
QY 861 TrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSer 880
Db 2983 TGGCTTCTGATGCGCTCTCTCAGTACACATCTATGGGTGACCCCTCCAGCCCTCAGC 3042
QY 881 HisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMet 900
Db 3043 CATTCAGGAACCGGAGCCGCCACCTAGCGGAGCGGCTGTGGCCCTCTCTGAAAGATG 3102
QY 901 GluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLys 920
Db 3103 GAAGGAACCACTCTCATCGGTACTCCAAGTCTTGAAGCCCACTTGGGAGACCCGAAA 3162
QY 921 ProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGlu 940
Db 3163 CCTCGGCCCTGCGAGCTGTCCACATTTGTCTATGGGCCCAAGCCACAGCCTTTGAATGAG 3222
QY 941 ThrAlaProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLysVal 960
Db 3223 ACGGCACCCAGTAATCTTTGGAACATCAGAAGCTGCTCTCCATTGACCTGCACAAAGTA 3282
QY 961 ValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSer 980
Db 3283 GTGTTACCAACTTTCGATCCCAATCGCCTCAAGTGAGACCTTTGCCCCCTGGAGAGGT 3342
QY 981 GlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGln 1000
Db 3343 GGTGCTGGAGCATCTCTGGAGGATCATGCTGGCGCGCTACAAACCAGAGGTGGGACAG 3402
QY 1001 ThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGln 1020
Db 3403 ACCATCCCGCTTCTCGCCTTCTCGGAGCATGTGTGGCCCTCTCTTGTGTGTACAG 3462
QY 1021 IleAsnLysAlaLysSerArgProLysArgTyrLysProArgValLysArgProGlnLeu 1040
Db 3463 ATCAGCAAGCCCAAGAGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3522
QY 1041 MetGlnGlnValHisProProLysThrProSerVal 1052
Db 3523 GCACAGCAGGCCCAACCCCTCGAAGGACCCCATCATG 3558

RESULT 9

BC026330

LOCUS

DEFINITION

Homo sapiens, similar to membrane-bound transcription factor

protease, site 1, clone MGC:26368 IMAGE:4816102, mRNA, complete

cds.

ACCESSION

BC026330

VERSION

BC026330.1

GI:20072867

KEYWORDS

MGC.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 3026)

Strausberg, R.

Direct Submission

Submitted (02-APR-2002) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

CDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki

Toshiyuki and Piero Carninci (RIKEN)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mod@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 32 Row: n Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA qi: 4506774.

```

FEATURES
Source
Location/Qualifiers
1 3026
/organism="Homo sapiens"
/db_xref="taxid:9606"
/clone="MGC:26368 IMAGE:4816102"
/tissue_type="Brain, hippocampus"
/clone_lib="NIH_MGC_95"
/lab_host="DH10B"
/note="Vector: pBluescript"
463 ..2121
/codon_start=1
/product="similar to membrane-bound transcription factor
protease, site 1"
/protein_id="AAH26330.1"
/db_xref="gi:20072868"
/translation="MKLVNIWLLLVLLGKKHGLDRLEKSKFEKAPGPCSHLTLLK
VESSVVEEYIVANGSYTAKARNSFTSSALKSEVDNNRIIPRNPSSDIPDSFE
VIOIKKQAGLLLEDPHNIKRVYKPRKLAIESDPTPCNRTGYSQKQWSSR
PLRSLRSGSGFPHATGRHSRLLRAIPROVAQTLOADVLMQYGTGACRVAVFD
TLSEKHPRFNKRNENTNERTLDGLGHGTFVAGVIASMECCQGPDAEHLHIFR
VFTNNGSVKVFEDAFNAILKLDVNLISGGDFMDHPFYDKQWELTANNVIVY
AICNGDPSGLTNNPADQYIGVGIGDFEDNIAFFSRGTTWELPGYGRMKPDLV
TIGNVPYSGYSGKCALSGTSVASVAGVATLIVSVQKRLVNPASMKQALIASA
RLPGVNNPFGECGKLDLLRAYQILNSYKQASLSPSIDITCEPTMMPYCSQPIYG
GMTVYNNVILNGGVGTGRIVDQPDWQYLPQNGDNIEVAFYS"
BASE COUNT. 778 a 707 c 797 g 744 t
ORIGIN
Alignment Scores:
Pred. No.: 2,47e-188 Length: 3026
Score: 3069.00 Matches: 580
Percent Similarity: 99.66% Conservative: 0
Best Local Similarity: 99.66% Mismatches: 2
Query Match: 54.64% Indels: 0
DB: 9 Gaps: 0
US-09-830-837-6 (1-1052) x BC026330 (1-3025)

```



```

Qy 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyr 500
Db 1903 CAGCAGAGTTGAGCCCGCCAGCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTAC 1962
Qy 501 CysSerGlnProIleTyrTyrGlyGlyMetProThrValValAsnValThrIleLeuAsn 520
Db 1963 TGTCTCCAGCCCATCTACTATGAGGAGGAAATCCGACAGTGTGTAATGTCACCATCTCCCAAC 2022
Qy 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGln 540
Db 2023 GGCATGGAGTCCACAGAGAAATGTAGATAAGCCTGACTGGCAGCCCTATTGTCACAG 2082
Qy 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGly 560
Db 2083 AAGCGAGACAACTGGAATGGCTTCTCTACTCTAGCTTATGCGCTTGTGTCGGC 2142
Qy 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGln 580
Db 2143 TACTTGCCATCTCCATTTCTGTGACCAAGAAAGCGGCTTCTTGGGAAGCATGTGCTCAG 2202
Qy 581 GlyHis 582
Db 2203 GGCCAT 2208

RESULT 10
LOCUS AK091212 3085 bp mRNA linear PRI 15-JUL-2002
DEFINITION Homo sapiens cDNA FLJ33893 fis, clone CTONG2007681, highly similar
to Site-1 protease of sterol regulatory element binding proteins.
ACCESSION AK091212
VERSION AK091212.1 GI:21749529
KEYWORDS oligo capping; fis (full insert sequence).
SOURCE Homo sapiens tongue, tumor tissue cDNA to mRNA, clone_lib:CTONG2
clone:CTONG2007681.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
AUTHORS Nishi,T., Ota,T., Nakagawa,S., Senoh,A., Mizuguchi,H., Inagaki,H.,
Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 3085)
Isogai,T. and Yamamoto,J.
Direct Submission
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: HRI and
RAB; annotation: HRI and RAB.
FEATURES
source
1..3085
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CTONG2007681"
/tissue_type="tongue, tumor tissue"
/clone_lib="CTONG2"
/notes="cloning vector: pME18SFL3"
BASE COUNT 760 a 763 c 790 g 752 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 6.19e-188 Length: 3085
Score: 3063.00 Matches: 569
Percent Similarity: 99.82% Conservativity: 0
Best Local Similarity: 99.82% Mismatches: 1
Query Match: 54.53% Indels: 1
DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x AK091212 (1-3085)

Qy 483 SerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSer 502
Db 691 AGTTTGAGCCCGCCAGCTACATAGATCTGACTGAGTGTCCCTACATGTGGCCCTACTGTCC 750
Qy 503 GlnProIleTyrTyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMet 522
Db 751 CAGCCCATCTACTATGAGGAGGATCCGACAGATGTTTAATGTCCACATCTCTCAACGGCATG 810
Qy 523 GlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGly 542
Db 811 GGAGTCACAGAGAAATGTAGATAAGCCTGACTGGCAGCCCTATTGTCACACAGCAGA 870
Qy 543 AspAsnIleGluValAlaPheSerTyrSerValLeuTrpProTrpSerGlyTyrLeu 562
Db 871 GACAACATTGAAGTTGCCCTTCTCTACTCTCGGTCTTATGGCCCTTGGTGGGCTACTGTG 930
Qy 563 AlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHis 582
Db 931 GCCATCTCCATTTCTGTGACCAAGAAAGCGGCTTCTTGGGAAGCATGTGCTCAGGGCCAT 990
Qy 583 ValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThr 602
Db 991 GTCATGATCACTGTGGCTTCCCGACAGACAGACAGTCAAAAAATGTTGCAGAACAGACT 1050
Qy 603 SerThrValLysLeuProIleLysValLysIleLeuProThrProProArgSerLysArg 622
Db 1051 TCACACAGT-AAGCTCCCATTAAGGTGAAGATAATCTCTACTCCCGCGGAGCAAGAGA 1109
Qy 623 ValLeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsn 642
Db 1110 GTTCTCTGGGATCAGTACCACACACCTCCGCTATCCACTGGCTATTTCCCGAGGATAT 1169
Qy 643 LeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPhArg 662
Db 1170 TTAAGATGAAGAATGACCTTCTAGACTGGAATGGTGTATGATCCATCCACACAAATTTCCAG 1229
Qy 663 AspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPhe 682
Db 1230 GATATGTACCAGCATCTGAGAGGATGGGCTACTTTGTAGAGGTCTCGGGGGCCCCCTTC 1289
Qy 683 ThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyr 702
Db 1290 ACCTGTTTTCATGCCAGTCACTAGTATGGACCTTCTGCTGATGTGGACAGTGGAGGAGTAC 1349
Qy 703 PheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIle 722
Db 1350 TTCCTCTGAAGAGATCGCAAGCTCCGAGGAGGAGTGGACAAACGGCTCTCGCTCGTCTATC 1409
Qy 723 PheSerAspTrpTrpAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
Db 1410 TTCAGTGACTGGTACAACTTCTGTTATGAGAAAGTGAAGTTTATGATGAAACACA 1469
Qy 743 ArgGlnTrpTrpMetProAspThrGlyIleValAlaAsnIleProAlaLeuAsnGluLeuLeu 762
Db 1470 AGCAGTGGTGGATGCCGATACCGGAGGAGCTAACATCCAGCTCTGATGATGAGTCTGTG 1529
Qy 763 SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsn 782
Db 1530 TCTGTGTGGAACATGGGTTTCAGCGATGGCTGTATGAAGGGGAGTTTACCCCTGGCCAAC 1589
Qy 783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVal 802
Db 1590 CATGACATGTATTATGCTCAGGTCAGCATCGCAAGTTTCCAGAAAGTGGCGTGTG 1649

```

QY	803	IleThrGlnThrPhelYsAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValVal	822
Db	1650	ATAACACAGAGCTTTCAAGGACCAAGGATGGAGTTTAAAGCAGGAGCAACAGCAGTGT	1709
QY	823	GluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleVal	842
Db	1710	GAACACGTCCTTATGGGACTTATCAGATGCCAGTCCAGCTGAGGCTGGAGCCGGATTGA	1769
QY	843	LeuTyrGlyAspSerAsnCysLeuAspSerHisArgGlnLysAspCysPheIleLeu	862
Db	1770	CATGATGGGACTCCATTCCTGGATGACATGCCACGACAGAGGACTGCTTTGGCTT	1829
QY	863	LeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSer	882
Db	1830	CTGATGCTTCCTCCAGTACATCTGATGGGTGACACCGCTTACGCTCAGTCACTCT	1889
QY	883	GlyAsnArgGlnArgProSerGlyAlaGlySerValThrProGluArgMetGluGly	902
Db	1890	GGAAACCCGACGCTCCAGTGGAGAGGCTGAGTCCAGTCCAGAGAGATGGAGGA	1949
QY	903	AsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArg	922
Db	1950	AACCATCTTCATCGGTACTCCAAAGTCTTGAGGCCCATTTGGAGACCCCAAACTCG	2009
QY	923	ProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsnGluThrAla	942
Db	2010	CTCTACAGAGCTGTCACGCTTCTTTGGGCCAAGCCACAGCTTTAAACGAGAGCGG	2069
QY	943	ProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysValLeu	962
Db	2070	CCAGTAACTTGGAAACATCAGAACTACTTCCATTTGACCTGGACAGAGTGGTGT	2129
QY	963	ProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSerGlyAla	982
Db	2130	CCCACTTTCGATCGATCGCTCAAGTGAAGGCTTGTCCCTTGGAGAGAGCGCGCC	2189
QY	983	TrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnValGlyGlnThrIle	1002
Db	2190	TGGACATCTCTGGAGGATCATGCTGGCGCTTACAAACGAGGAGTGGCGAGACCAT	2249
QY	1003	ProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGlnIleAsn	1022
Db	2250	CCGCTTTTGGCTTCCTGGAGGCCATGGTCTGCTGCTTCTTGTGTACAAATCAAC	2309
QY	1023	LysAlaLysSerArgProLysArgArgLysProArgLysValLysArgProGlnLeuMetGln	1042
Db	2310	AGGCCAAGAGAGCGCCGAAGCGAGGAGCCCAAGGAGGAGCCGAGGAGCCGAGCAG	2369
QY	1043	GlnValHisProProLysThrProSerVal	1052
Db	2370	CAGGTTCAACCCGCAAGACCCCTTGGGTG	2399
RESULT 11			
AF441758			
LOCUS	AF441758	3138 bp	linear
DEFINITION	Drosophila melanogaster serine endopeptidase mRNA, complete cds.		
ACCESSION	AF441758		
VERSION	AF441758.1	GI:20378354	
KEYWORDS			
SOURCE	Drosophila melanogaster.		
ORGANISM	Drosophila melanogaster.		
REFERENCE	Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.		
AUTHORS	Seegmiller, A.C., Dobrosotskaya, I., Goldstein, J.L., Ho, Y.K., Brown, M.S. and Rawson, R.B.		
TITLE	The SREAP pathway in Drosophila: regulation by palmitate, not sterols		
JOURNAL	Dev. Cell 2 (2), 229-238 (2002)		
MEDLINE	21822137		
PUBMED	11832248		
REFERENCE	2 (bases 1 to 3138)		
AUTHORS	Seegmiller, A.C., Dobrosotskaya, I., Goldstein, J.L., Brown, M.S. and Rawson, R.B.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-NOV-2001) Molecular Genetics, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9046, USA		
FEATURES	Location/Qualifiers		
source	1..3138		
	/organism="Drosophila melanogaster"		
	/strain="Oregon R"		
	/db_xref="taxon:7227"		
	/chromosome="3"		
	/map="3L"		
	/dev_stage="embryo"		
	58..3096		
	/note="dsIP; similar to CP22149"		
	/codon_start=1		
	/product="serine endopeptidase"		
	/protein_id="AAM20922.1"		
	/db_xref="GI:20378355"		
	/translation="MNVTPLFIISATCSLDFAKTAVPVNEFIVHPSKVFAPVRESY IAKLLGSNTWRIYVRLNLAQPSDFDLKVCDSYESSSEFIIBRLQTHPSKAV VPORSRRILANTYDAYSLTYIHRHPQVLNPNDRHROLCSVLHNLKWLKGTG KGVVAIFDTGTHKHPFRVKNRTWTKSLDDRVSHGTFVAGVIASSRCLGFA DDADLIKFTFTNSOVSYSFLDAFNAYIRKINLNLSDPDMSPFVSKVLEL SANNVIMISAGNDGQLXGTLNPGSDVGVGGIOPDKIAKFSRGTWELPLG YKNGLDITVYGSQVSDVEKGRISCTSVSSPVVAGAAALISGAFKIDYINPA SLKQVLEGRKLRHYNMFGKGLNLSKMLLSKPKITLIPALVLFDTONYMP YSOPLITGSSVALIANVILNGISVSHVIGPKITPROFNOGLOVSAQVSPVMP WTGMSVFIANKRGNEFGVSGSITLVLESEKOTNETHVTEVDFPLIKVTPKPP RNKRIIDQTHSLRYPRIIPRODLKVLQDLPDWRADNHIHNFROMKYTHLRNYYVID VLEPFCTFNASDILVIDPFGDEINAIQDKNYKGLNIVFGVGYNTYMK KIKFPENQDITVTGTGANIPALNDLLKPGIAPDFVGGHFKLGDMSMYASGA TIYKFPMPGDIIIVTKLNDGSLINSKTPSKVAKLDVIFGMFANRANSQSNSEI VYNAESNLAEIPTDYTFKRNKLVLLKORSISFAKSNHNETRKNRIAYGSIKSL DSHLEKACTWLLITFLDFAINSLKSLQNLNRIETFKHETRAPLIRISQSIKSR SDNCEQFKWLAPTKONNAERKSSIDVTILENEHEHINILKNLLEETAKLGONN DYLTGMSADSLTPFIILISLVILMFLVFLKRRFIL"		
BASE COUNT	981 a	662 c	656 g
ORIGIN	839 t		
Alignment Scores:			
Pred. No.:	3.7e-144	Length:	3138
Score:	2383.50	Matches:	478
Percent Similarity:	59.83%	Conservative:	167
Best Local Similarity:	44.34%	Mismatches:	272
Query Match:	42.43%	Indels:	161
Db:	3	Gaps:	17
US-09-830-837-6 (1-1052) x AF441758 (1-3138)			
QY	47	PheSerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAla	66
Db	112	TTTAAACAGCCGCTGTTCCTCAATGAGTTCATCTTCTTCAATCAATACTTGGC	171
QY	67	LysAlaArgAsnSerPheIleSerSerAlaLeuLysSerGluValAspAsnTrpArg	86
Db	172	CCGGTCCGAGAACTCTACATCGCAGCAAACTTCTGGTTCAACCTACGACATGGAGA	231
QY	87	IleIleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIle	106
Db	232	ATTGTTCCTCCCTTAAATTTGGCTGGCAATATCAAGTATTTGATATCTTACAGTT	291
QY	107	LysGluLysGlnLysAlaGly-----LeuLeuThrLeuGluAspHisProAsn	122
Db	292	TGGCAGCGTATGAATCATCATCATCATATATATATAAGAGCGCTTCAGACTCACCA	351
QY	123	IleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyr---AlaGluSer	141
Db	352	GTAAAGCAGTAGTTCCTCCCGCAGCGGTACCAAGAGTCTCTAAACTATGACCGCTAT	411
QY	142	AspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerArgPro	161


```
Db 412 AACCTAACG----- 420
Qy 162 LeuArgAlaSerLeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArgHisSer 181
Db 421 -----TATATTCAC-----CGGCATCCC 438
Qy 182 SerArgArgLeuLeuArgAlaIlePro-----ArgGlnValAlaGlnThr 196
Db 439 CAAGGAGTGTAAAGGAACAGAAACCAACACGATCGCCACCGACAAATGTGTCGGTA 498
Qy 197 LeuGlnAlaSerValLeuTrpGlnMetGlyThrThrGlyAlaAsnValArgValAlaVal 216
Db 499 CTCACGCCAACATCTTTGGAAGCTGGGTATCATCAGGCAAGGAGTAAAGTGGCCATT 558
Qy 217 PheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsn 236
Db 559 TTCGACACTGGCCTAACCAAAACCATCCACACTTTCGAANTGTAAAGGAACGAACAAC 618
Qy 237 TrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyVal 256
Db 619 TGGACGAATGAAAGTCACTTGACGACAGAGTCAGTCATGCGACCTTCGTCGCGGGGTA 678
Qy 257 IleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArg 276
Db 679 ATCGCTTCTCCAGGGAATGCTTAGGCTTCGTCGCCGACCGCGATCTTTCATATTTAAA 738
Qy 277 ValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAla 296
Db 739 GTTTTACGAACCTCCCAAGTTCTTACACTTCTGTTCTGTTCTGATGATCACTACGCG 798
Qy 297 IleLeuLysIleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHis 316
Db 799 ATATATAGGAAATAAACAATCTCAACCTTAGCATTTGGGGTCCCGACTTATGGACTCG 858
Qy 317 ProPheValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIle 336
Db 859 CCGTTCGTGAAAGGTGTGGAACTGTGCGCTAATAATGTCAATGATATGGCAGCA 918
Qy 337 GlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIle 356
Db 919 GGAATGATGTCCTTGTACGGCCACGCTAAACAATCTCGCGATCAGAGCGATGTAGTT 978
Qy 357 GlyValGlyIleAspPheGluAspAsnIleAlaArgPheSerArgGlyMetThr 376
Db 979 GCGCTTGGTGCACTTCAGTTGTATGATAAATCCCAAGTTTACTTCAGAGGAATGACA 1038
Qy 377 ThrTrpGluLeuProGlyGlyThrGlyArgMetLysProAspIleValThrTyrGlyAla 396
Db 1039 ACGTGGAACTTCCCTTAGGCTACGGAGTATGGGACTCGATATGTCACGTACCGAAGT 1098
Qy 397 GlyValArgGlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAla 416
Db 1099 CAAGTGAAGCAGTGTATGCGCAAGGGGTGACAGCAGCTCTCTGGAACATCCGTGCC 1158
Qy 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435
Db 1159 TCTCCAGTTGTGAGGGGTGCTGTCACCTGCTGTATAAGCGGTGCTATTCAGAAATCGAC 1218
Qy 436 LeuValAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGly 455
Db 1219 TACATAACCCAGCATCTCTTAAGCAGGTACTTCAATGAAGTGGCCGAGAACTCCCGCAT 1278
Qy 456 ValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeu 475
Db 1279 TATAACATGTTTGACAGGGAGCTGGAAACTGAATTTGCTGAAGAGTATGCACTATTG 1338
Qy 476 AsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysPro 495
Db 1339 CTGTATACAAACCAAGATAAACCTTATTCGCGCATACCTTGACTTTCACCCAA---AAC 1395
Qy 496 TyrMetTrpProTyrCysSerGlnProIleTyrGlyMetProThrValValAsn 515
Db 1396 TATATGGGCTTATAGCTCCCAACCTCTGTACTTGAAGCTCCGTCGCTATTGCAAC 1455
```

```
Qy 516 ValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGln 535
Db 1456 GTTACCACTACTCAATGATCTCTGTCACAGTCATATAGTTGGCATCCCTTAATGGATT 1515
Qy 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeu 555
Db 1516 CCCGATTTCGAAAACCAAGCTCAGTTCTTCAAGTATCTGCACAAGTTTCCGCTATCGTT 1575
Qy 556 TrpProTrpSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTrp 575
Db 1576 TGGCCGGGAGCGGTGGATGTCAGTTTATTCTCTGTAATAAAGGAAGAGAGAATCTTT 1635
Qy 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595
Db 1636 GAAGGTCTTGTAAAGGAAGTATCACCTAGTTTGGAAAGCTTTAAACAGACACCAAC 1695
Qy 596 LysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIleIlePro 615
Db 1696 GAAACT-----CATGTTACAGAAGTCGACTTCTCTTAAACAATAAAGGTTACTCCA 1746
Qy 616 ThrProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyrProPro 635
Db 1747 AAACCCCAAGAAACAAGAGATTATGGATCAGTACCACACCTTAAGATATCCACCG 1806
Qy 636 GlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp 655
Db 1807 CGCTATATTCACGAGATGATCAAAAGTTAAACTAGATCCTCTGGACTGGAGGCAGAC 1866
Qy 656 HistIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal 675
Db 1867 CATATACACAAACTTTAGGACATGATATACACATTTACGAAATGTTGGCTACTACATT 1926
Qy 676 GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet 695
Db 1927 GATGTTTGGCAGAACCTTCACTGCTTCAATGCCCGGATTTAGCGGCTTATGATT 1986
Qy 696 ValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgAspVal--- 714
Db 1987 GTTGACCTGAGAGAGGTTTGGCGACGAGGAAATAAAGCTTTTACAGGAAACGTTGAT 2046
Qy 715 AspAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerValMetArgLys 734
Db 2047 AAAAGAGGCTTGAATGCTGCTGATTCGGAGCTGGTATACACACTGTGTGTAATAA 2106
Qy 735 ValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAlaAsn 754
Db 2107 ATTAATTCCTTGACGAGACACCCGACATGGTGGACACCCGACACTGGTGGCGCAAT 2166
Qy 755 IleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyr 774
Db 2167 ATTCAGCTTGAATGATTATTGAAGCCATTGGAATTTGCTTTTGGCGATTGTCGCT 2226
Qy 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla 794
Db 2227 GAGGACATTTCAAACTGGCGACCATTCATCTACTATGCTAGTAGGACCCACATTTGT 2286
Qy 795 LysPheProGluAsp-----GlyValValIleThrGlnThrPheLysAspGlnGlyLeu 812
Db 2287 AAGTTTCAATGAATCCAGGAGATATTATAGTGGGCACAAACCTGAATGACCAAGACTT 2346
Qy 813 GluValLeuLysGlnGluThr-----AlaValValGluAsnValProIleLeuGly 829
Db 2347 TCGATTATTATTTCTAAACACACCCAGCAAGGTAGCAAAACCTAGATGTACCTATTTTGGT 2406
Qy 830 LeuTyrGlnIleProAla----- 835
Db 2407 ATGTTCCAAACCAAGCGCAACAGTATTCAAAGCAACGAGGAATCGTGGTCAATGCGGAA 2466
Qy 835 ----- 835
Db 2467 AGCAATTTGGCAGAGGCTATATCCACAGATTACTCCACATTTAAGAACCAGGTTTTCCTA 2526
```



```
QY 277 lphetrhAsn----- 280
Db 53985 TTTTACGACCTCCAGAGTGGCAACAGCTAGAACCTGAATAACTAAGCCAACTAATTAG 54044
QY 281 -----Asn-GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleL 298
Db 54045 CCTACATCGCGCAGGTTTCTTACACTCTCGTGGTTCCTGGATGCATTCACATAGCGCATAT 54104
QY 298 eulysIysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProp 318
Db 54105 ATAGGAAATAAACAATCTCAACCTTAGCATGGGGTCCCGACTTTATGGACTCGCCCT 54164
QY 318 heValAspIysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyA 338
Db 54165 TCGTTGAAAGGTGTGGAACTGTGCGCTAATAATGATCATATGATATCGGAGCAGGAA 54224
QY 338 snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV 358
Db 54225 ATGATGGTCCCTGTAGCGCAGCTTAACATCTCTGGCATCAGAGCGATGTAGTTGGCG 54284
QY 358 alGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThr 378
Db 54285 TTGGTGGCATTCAGTTTGATGATAAAATCGCCAGTTTAGTTCGGAGAGATGACACGT 54344
QY 378 rpGluLeuProGlyGlyTyrGlyArgMetIysProAspIleValThrTyrGlyAlaGlyV 398
Db 54345 GGAATCTCCCTTAGGCTACGACGTATGGACTTCGATATTCACGTACGGAAGTCAAG 54404
QY 398 alArgGlySerGlyValIysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerP 418
Db 54405 TGAAGGAGAGTGTGTCGCGAAGGGTGCAGACGACTCTCTGGAACTATCGTCTCTCTC 54464
QY 418 roValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnIysArgGluLeuV 437
Db 54465 CAGTTGTTCAGGGGTGTGCTGCTGCTTATAAGCGGTGCATTTCAAGAAATCGACTACA 54524
QY 437 alAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValA 457
Db 54525 TAAACCCAGCATCTTTAAGCAGGTACTCATTTGAAGTGCAGAGAACTGCCCATATA 54584
QY 457 snMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnS 477
Db 54585 ACATGTTTGACAGGAGGTGGAAACTGAATTGCTGAAGAGTATGCAGTATTCCTCT 54644
QY 477 erTyrIysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM 497
Db 54645 CATACAAACCAAGATAACCTTTATTCGGCATACCTTCCTCCTACCCAA---AACTATA 54701
QY 497 etTrpProTyrCysSerGlnProIleTyrGlyGlyMetProThrValValAsnValT 517
Db 54702 TGTGGCCTTATAGCTCCAACTCTGTACTATGGAAGTCCGCTGCTATTCGAACGTTA 54761
QY 517 hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspIysProAspTrpGlnProT 537
Db 54762 CCNACTCAATGGTATCTCTGTCAACAGTCATATAGTTGGCATCCCTAAATGATTCGCG 54821
QY 537 yrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeuTpp 557
Db 54822 ATTTCGAAACCAAGTCAAGTTCTTCAAGTATCTGCACAAGTTTCGCTCATCGTTTGC 54881
QY 557 roTrpSerGlyTyrLeuAlaIleSerIle----- 566
Db 54882 CGTGGACCGGTGGATGTCAGTTTATTATGTCAGTATGAAATAGCAACGCTATAAA 54941
QY 566 ----- 566
Db 54942 AGGATATATGGGTACATATTGATATGAAAGTCTAAGCCATTCTATACCAACAATTTT 55001
QY 567 -----Ser-ValThrLysIysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal 583
Db 55002 TGTWTACACGCTGTAAAAAGGAGGAGAAACTTTGAAGGTGTTTGAAGGAGTATC 55061
QY 584 MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer 603
```

```
Db 55062 ACCTAGTGTGGAAAGCTTTTAAACAGACACCACCAACAACT-----CATGTTACA 55112
QY 604 ThrValLysLeuProIleLysValLysIleIleLeuProThrProProArgSerLysArgVal 623
Db 55113 GAAGTCGACTTTCCTTTAAACAATAAAGGTACTCTCAAAACCGCCCAAGAAACAAGAGTAT 55172
QY 624 LeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLeu 643
Db 55173 TTATGGATCAGTACACACAGCCTTAAGGTATCCACGCGCTATATTCACAGAGATGATCTC 55232
QY 644 ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAsp 663
Db 55233 AAGTTAAACTAGATCTCTGGACTGGAGGCGACACCATATACACAAACTTTAGGGAC 55292
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
Db 55293 ATGTATACACATTTACGAAATGTTGCTACTACTACATGATGTTTGGGAGAACCTTCACC 55352
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuMetValAspSerGluGluTyrPhe 703
Db 55353 TGTCTCAATGCTCGGATATGCGGTTATGATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 55412
QY 704 ProGluIleAlaLysLeuArgArgAspVal---AspAsnGlyLeuSerLeuValIle 722
Db 55413 GACGAGGAAATAAACGCTTTACAGGAAACGCTGATATAAGAGGCTTGAATCTGCTGCTGA 55472
QY 723 PheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr 742
Db 55473 TTCGAGACTGCTATACACACCTGCTGATGAAAAAATAAATCTTTTACGAGAACACC 55532
QY 743 ArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeu 762
Db 55533 CGACAATGCTGGACACCCGACACTGCTGGCGCAATATTCACGCTTGAATCAATTTATG 55592
QY 763 SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsn 782
Db 55593 AAGCCATTGGAATTTGCTGCGATTTTGTGCGTGGGACATTTTCAAACTTTCAGCGCAG 55652
QY 783 HisAspMetTyrTrpAlaSerLysSerIleAlaLysPheProGluVal-----Gly 800
Db 55653 CATTCATGTACTATGCTAGTGGAGCCACATTTAAAGTTTCCATGATGATCCAGAGAT 55712
QY 801 ValValIleThrGlnThrPheLysAspGlnGlyLeuVal----- 814
Db 55713 ATATAGTGGCGCAAAACTGAATGCCAGGACTTTCGGTAATTTAATATATATATATAT 55772
QY 815 -----LeuLysGlnGluThr 819
Db 55773 TTAGTTTCCACAGAAATACAACTTTCCTTCCAGATATTAATTTCAAAACACC-AGC 55831
QY 820 AlaValValGlu---AsnValProIleLeuGlyLeuTyrGlnIleProAla----- 835
Db 55832 AAGGTAGCAAACTAGATGATGATCTATTTTGGTATGTATTCCAAACCAAGCGCAACAGTAT 55891
QY 835 ----- 835
Db 55892 CAAGCAACGAGGAAATCTGCTCAATCGGAAAGCAATTTTGGCAGGCTATATACCCACA 55951
QY 835 ----- 835
Db 55952 GATTACTCCACATTTAAGAACCGGTTTGTCTGCTGCGAAACGAGCAACAGTATCAGT 56011
QY 836 -----GluGlyGlyValGlyIleValLeuTyrGlyAsp 846
Db 56012 TTTGCAAAACCAATAATCATGAACTAAGAATCAAGGAGTATGCTGCTATATGCGGAC 56071
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAlaLeu 866
Db 56072 TCCACTGCTCCGACTCCACGATCTGGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 56131
QY 867 LeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGln 886
```

```
Db 56132 TTAGATTTCATAAAC-----TCGCAC----- 56155
QY 887 ArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHis 906
Db 56156 -----AAATCAAGTTTATTGCAG 56173
QY 907 ArgTySerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
Db 56174 AATCTAAATCGTATACTGAATTCACAAATAGAGAGACCACTACCCCTTAGGATA 56233
QY 927 -----CysProargLeuSerTTPala 933
Db 56234 TCACAAGATTTATAAAATCTCGTTCACAGGACAATAATGTGAACAATTTAAGTGGCTT 56293
QY 934 LysProGlnProLeuAsnGluThr-AlaProSerAsnLeu----- 946
Db 56294 GCACCCAGCAGCAAAATACGCCGAGGAAGAAATCTCTATATAATAGACGTAAACATA 56353
QY 947 -TrpLysHisGlnLysLeuSerLeuAspLeuAspLysValValLeuProAsnPheAr 966
Db 56354 CTGGAATAATGAAGGTATTTATATACATTTAGTTCAACTACAAAGTCTCATATTTGTTGT 56413
QY 966 gSerAsnArgPro-----GlnValArgProLeuSerProGlyGluSerGlyAlaTrpAs 984
Db 56414 TTCAGAACACGAGATAAATTTATCAAAATTTATTGGGTGAGGAGATCGCAAACTAGG 56473
QY 984 pIleProGlyGlyLeuMetProGlyArgTyAsnGlnGluValGlyGlnThrIleProVa 1004
Db 56474 GCAAACATGATTATTACAGGAATGCATCCGGGATGCTTAATGACTCCAATAT 56533
QY 1004 lPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValGlnIleAsnLys 1023
Db 56534 ATTCTTAATATATTAAGCTAATTTCTATCATGTTATTGTTTATTTTAAAGCGT 56591

RESULT 13
AE003595/c
LOCUS 294308 bp DNA linear INV 04-OCT-2000
DEFINITION Drosophila melanogaster genomic scaffold 142000013386036 section 5
of 9, complete sequence.
ACCESSION AE003595
VERSION AE002647
KEYWORDS HTG.
SOURCE GI:10726941
ORGANISM Drosophila melanogaster.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 294308)
Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.F., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazee,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferrieria,S.,
Fleischmann,W., Fowler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,X., Levitsky,A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattel,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
```

```
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shner,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
MEDLINE 10731132
PUBMED
REFERENCE 2 (bases 1 to 294308)
AUTHORS Adams,M.D., Celnikier,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7296463.
FEATURES
source 1. 294308
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
/complement(<322..>837)
/gene="CG14574"
/evidence=not_experimental
/map="78E2-78E4"
/db_xref="FLYBASE:FBgn0014574"
/db_xref="FLYBASE:FBgn0037104"
/complement(<322..>837)
/gene="CG14574"
/product="CT34306"
/evidence=not_experimental
/db_xref="FLYBASE:FBgn0014574"
/db_xref="FLYBASE:FBgn0037104"
/complement(322..837)
/gene="CG14574"
/note="CG14574 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF51750.1"
/db_xref="GI:7296464"
/db_xref="FLYBASE:FBgn0014574"
/db_xref="FLYBASE:FBgn0037104"
/translation="MQPTLECSILGSWYASEIPNVFDLNGTDDAGGAAAAALSCS
RYPRIWATSPCOTQRLCRGPPPAHLQPHLLLVRGEGDGPGRCTTWRDWSGRL
DSNGKDKQRGGGTSTERACPCDISGLPTGHNCNPNRDLDKAPGPGTQSGHSTH
NKPSSPSQAR"
<27608..>32838
/gene="AP-2"
/note="AP-2"
/map="78E4-78E5"
/db_xref="FLYBASE:FBgn0007807"
/db_xref="FLYBASE:FBgn0023417"
/join(27608..27865,28904..29134,29204..29371,30628..30835,
31480..31669,31730..31994,32098..32265,32329..32838)
/gene="AP-2"
/product="CT32720"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."
/db_xref="FLYBASE:FBgn0007807"
/db_xref="FLYBASE:FBgn0023417"
/join(27821..27865,28904..29134,29204..29371,30628..30835,
31480..31669,31730..31994,32098..32265,32329..32439)
/gene="AP-2"
/note="AP-2 gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
```



```
Db 36171 | |||:|||||:|||||:||||| ||||| |||:|||||:||||| 36112
Qy eAspThrGlySerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTr 237
Db 36111 CGACACGGCTACCAACCAACCATCCACACCTTCGGAATGTAAGGAAGCAACAACG 36052
Qy pThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValII 257
Db 36051 GACGAATGAAGTCACCTTGACGACAGAGTCAGTCATGGCCACCTTCGTCGCCGGGTAA 35992
Qy eAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisPheArgVa 277
Db 35991 CGCTTCTCCAGGGAATGCCCTAGCTTCGCTCCGACGCCGATCTTTACATAATTTAAAGT 35932
Qy lPheThrAsn- - - - - 280
Db 35931 TTTTACGAAGCTCCCAAGTAGCAAGCTAGAACTGAATTAACCTAAGCAATCAATTAG 35872
Qy 281 - - - - - Asn-GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleL 298
Db 35871 CCCTACATCGGCAAGTTCTTACATCTCTGGTTCCTGGATGCATTCACATACCGGATAT 35812
Qy euLysLysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProp 318
Db 35811 ATAGGAAATTAACATCTCAACCTTAGCATTTGGGGTCCCGACTTTATGGACTCGCGCT 35752
Qy heValAspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyA 338
Db 35751 TCCTTGAAAGGTGTGGAACTGTCGCTAATATGTCATTAATGATGATGGCAGAGCAAA 35692
Qy snAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyV 358
Db 35691 ATGATGGTCCCTTGTCGGCCACGCTAAACAATCTGCGCATCAGACGGATGATGGCG 35632
Qy 358 alGlyGlyIleAspPheGluAsnIleAlaArgPheSerSerArgGlyMetThrThrT 378
Db 35631 TTGGTGGCATTCAGTTTGATGATGAATAATCCCAAGTTTATGTCAGAGGAATGACAACGT 35572
Qy rpGluLeuProGlyGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyV 398
Db 35571 GGAACCTCCCTTAGGCTACGAGCATGATGGGACTCGATATGTGACGTACGGAAGTCAAG 35512
Qy 398 alArgLysGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerP 418
Db 35511 TGAAGGCACTGATGTGCGCAAGGGTGCAGACACTCTCTGGAACATCGGTCTCTCTC 35452
Qy 418 roValValAlaGlyAlaValThrLeuLeuValSer- - - ThrValGlnLysArgGluLeuV 437
Db 35451 CAGTTGTTCAGGGGCTGCTGCACTGCTTATAAGCGGTGCATTTTCAGAAAATCGACTACA 35392
Qy 437 alAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgArgLeuProGlyValA 457
Db 35391 TAAACCCAGCATCTCTTAGCAGGTACTCATTTGAAGGTCCGAGAACTGCCCATATATA 35332
Qy 457 snMetPheGluGlnGlyHisGlyLysLeuAspLeuArgAlaTyrGlnIleLeuAsnS 477
Db 35331 ACATGTTTGAGCAGGAGCTGGAATACTGAATTTGCTGAAGACTATGCAGCTATTGCTGT 35272
Qy 477 erTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrM 497
Db 35271 CATCAAAACCAAGATAACCTTATTTCGCGCATACCTTGACTTCACTTCAACCCAA- - - AACATA 35215
Qy 497 etTrpProTyrCysSerGlnProIleTyrTyrGlyGlyMetProThrValValAsnValT 517
Db 35214 TGTGGCTTATAGCTCCCAACCTCTGTACTATGGAAGCTCCGTCGCTATTGCAACAGCTTA 35155
Qy 517 hrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnProT 537
Db 35154 CCATACCTCAATGCTATCTCTGTCACAAAGTCATATAGTTGGCATCCCTCAATTAATGGAATCCCG 35095
Qy 537 yrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrpP 557
| | | | | : : : : : | | | | |
```

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu

MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.

NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is T020D01; 3' clone is F028I16. Actual start of this
clone is at base position 1 of T29J13; actual end is at 100592 of
T29J13.

FEATURES

source
1. 100592
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="V"
/map="unknown"
/clone="T29J13"

BASE COUNT 30991 a 18500 c 18948 g 32153 t
ORIGIN

Alignment Scores:
Pred. No.: 3,82e-115 Length: 100592
Score: 1965.50 Matches: 447
Percent Similarity: 49.09% Conservative: 149
Best Local Similarity: 36.82% Mismatches: 255
Query Match: 34.99% Indels: 368
DB: 8 Gaps: 23

US-09-830-837-6 (1-1052) x T29J13 (1-100592)

QY 48 SerSerThrValValGluTyrGluTyrIleValAlaPheAsnGlyTyrPheThrAlaLys 67
DB 35394 TCTGAGATCAAAACGAAGACTAATCTATATATCGATTAAAGCAGTATAGCCGCGAAA 35335
QY 68 AlaArgAsnSerPheIleSerSerAlaLeuLysSerSerGluValAspAsnTrpArgIle 87
DB 35334 GATCATCGAATTTACCTCGAATCGAAGCTCGATCCGCT-----GGTTGGGTTGG 35284
QY 88 IleProArgAsnAsnProSerSerAspTyrProSerAspPheGluValIleGlnIleLys 107
DB 35283 ATCGAGAGATTAAACCGCGGACGAAGTATCCGACTGCTCGGTGTTCTATGATCGAG 35224
QY 108 GluLysGlnLysAlaGlyLeuLeu---ThrLeuGluAspHisProAsnIleLysArgVal 126
DB 35223 GAACCGGGAAGAGCCCTGGTTGAGAGATTGAGAGCTGGAGATGGTGAACAGCTG 35164
QY 127 ThrProGln----- 129
DB 35163 AACGTTGAGTTTAAGTACCAGAGAGCTTTTTCCTCGAGAGATCTTTTCCCGACGGGAGAAA 35104

QY 835 ----- 835
DB 33964 GATTACTCCACATTAAAGACCGGTTTGTCTACTCGCAACGAACGAAGTATCACT 33905
QY 836 -----GluGlyGlyValValGluValLeuValLeuValGlyAsp 846
DB 33904 TTTGCGAAAGCAATAATCATGAAACTAAGAACTAAGAGACGTTATGCGGTATATGGGAC 33845
QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
DB 33844 TCCAACTGCTCGACTCCACGCACTCGAGAGCGTTGCTACTGCTGCTAATAACGTTT 33785
QY 867 LeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSerGlyAsnArgGln 886
DB 33784 TTAGATTTTGCATAAAC-----TCGCAC----- 33761
QY 887 ArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHis 906
DB 33760 -----AAATCAAGTTATTGCGAG 33743
QY 907 ArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla 926
DB 33742 AATCTAAATCGTATACTGAATTCACAAATTAGAGAGACCACTACCCCTTAGGATA 33683
QY 927 -----CysProArgLeuSerTrpAla 933
DB 33682 TCGCAAGTATTATAAAATCTCGTTACAGGACAATAATTGTGAACAATAATTAAGTGGCTT 33623
QY 934 LysProGlnProLeuAsnGluThr-AlaProSerAsnLeu----- 946
DB 33622 GCACCGGAGCAAAATAACCGGAGAAAGAAATCTCTATAATAGACGTAACCATTA 33563
QY 947 TrpLysHisGlnLysLeuSerIleAspLeuAspLysValValLeuProAsnPheArg 966
DB 33562 CTGGAATAAGAGTATTATATACATTAGTCAACATCAAACTCATATTGTTGTT 33503
QY 966 gSerAsnArgPro-----GlnValArgProLeuSerProGlyGluSerGlyAlaTrpAs 984
DB 33502 TTCAGACACAGCAATAATTTAATCAAAATTTTGGTGGTGGAGGATCGCAAACTAGG 33443
QY 984 pileProGlyIleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIleProVa 1004
DB 33442 GCAAACAACTATTATTACAGGAGTGCATCCGGATAGTCTAATGACTCCAAATTAT 33383
QY 1004 lPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGlnIleAsnLys 1023
DB 33382 ATCTTAATTAATTAAGCTAATGTTATCATGTTATTGTTTATTATTTAAAGCGT 33325
RESULT 14
T29J13/c 100592 bp DNA linear PLN 23-AUG-2000
LOCUS Arabidopsis thaliana BAC T29J13.
DEFINITION AF296838
ACCESSION AF296838.1 GI:9885851
VERSION
KEYWORDS
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 100592)
AUTHORS Latreil, P., Courtney, L., Abbott, A., Gregory, S. and Bielicki, L.
TITLE The sequence of A. thaliana T29J13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 100592)
AUTHORS Washington University Genome Sequencing Center.
TITLE The A. thaliana Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 100592)
AUTHORS Wilson, R.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-2000) Department of Genetics, Washington

QY 130 -----ArgLysValPheArgSerLeuLysTyrAlaGluSerAspProThrValProCys 147
|||:|||||
Db 35103 CGTCCGGGGAAGATCTTCTATGCTCTTGAAGAAGAACTCAGTCTCTCTATG 35044
QY 148 AsnGluThr-----ArgTrpSerGlnLysTrpGlnSerSerArgProLeu 162
:|||||
Db 35043 GCGGATACAAAGTAACACCAGGTAAACTGGAGTCGACA-----TCCTCTCGCTCAGGT 34991
QY 163 ArgArg-----AlaSerArgSerLeuGly 170
:|||||
Db 34990 GAGAAATGTGAAGTAACATAGTCAATATCTGTTTCTCTGCTCACTATCGATCTATGT 34931
QY 171 SerGlyPheTrpHisAlaThrGlyArgHisSerSerArgLeu-----LeuArgAlaIle 189
:|||||
Db 34930 GCTAATTTTGGTAT-----CGTAAACTGTAAGTGAGGTCTTTA 34892
QY 190 ProArg-----GlnValAlaGlnThrLeuGlnAla 199
:|||||
Db 34891 AT-AAGTTACATTTGTTTCTATTTGCGAAGACTCAAGTCAAGTCCATGTTTGGAGCA 34833
QY 200 AspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValPheAspThr 219
|||:|||||
Db 34832 GATCATCTCTGGAAGAAGGTACACCGGCGCTAAAGTTAAAGTGGCTATATTTGATACT 34773
QY 220 GlyLeuSerGlnLysHisProHisPheLys----- 229
|||:|||||
Db 34772 GGAATAAGAGCTGATCACCTCATTTTCG-TAAGATCAAGGTATCACTTTTGGAGAAAC 34714
QY 229 ----- 229
Db 34713 TGATAAACCATAGTATTCTCAAGATTTCTGAGAAGTGTGAAGTGTAAATTAATAGT 34654
QY 230 -----AsnValLys 232
|||:|||||
Db 34653 TTTCCTTTTATCTGCTTGTAGCTCTGAAGTGTATTCGACGGAGTAACCTTCAG 34594
QY 233 GluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPhe 252
|||:|||||
Db 34593 GAGCGTACCAATTTGGACAAATGAGGACACTTTGAATGACAACCTGGTGTATGGAGCTTT 34534
QY 253 ValAlaGlyValIleAlaSerMetArg--GluCysGlnGlyPheAlaProAspAlaGlu 271
|||:|||||
Db 34533 GTTCTGGGTGTTATCTGCTGCGAAATCCAGAAATGCTTGGTTTTCCTGCTCAGACACGGAA 34474
QY 272 LeuHisIlePheArgValPheThrAsn-Asn----- 281
:|||||
Db 34473 ATCTATGCTTTTCGAGTGTTCACAGATGCCAGGTAATGCATGATCACACAACAGATACT 34414
QY 281 ----- 281
Db 34413 TCTTCTCATCGCGGATAAGGTTTAACATTCTTCAACTAACTAAATCCACTCATGGTCC 34354
QY 281 ----- 281
Db 34353 AGTAAATCTGTTCTGTTTCCCTTAATACACTTGTGAAGTACTGATATTTGTTACACA 34294
QY 281 ----- 281
Db 34293 GTGATCTTTATTTGTCCTCGAGACACTGTAATTTGAGCGCTATAAATCTGTGTCCTCC 34234
QY 281 ----- 281
Db 34233 TTTTCTCTTTCCAGATACGCAAGTTGCTTAATCAATAAGGTTATGATAAGTTTCGCT 34174
QY 282 -----GlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLe 298
|||:|||||
Db 34173 CAGCTTGTTCAGGATATACACCTCATGTTTCTGTGAGCTTCAATATGCTATAGC 34114
QY 298 uLysLysIleAspValLeuAsnLeuSerIleGlyLysProAspPheMetAspHisProPh 318
:|||||
Db 34113 AACAGATATGGATGATTAATTTAGCATTTGGGGACCGGATTAATCTGGAATCTGCTTT 34054
QY 318 eValAspLys----- 321

Db 34053 TGTGAGAAAGGTAATTTGTTCTTAATCTCTTGGTGGAGTTTACTTATCTCTGTGAGTCG 33994
QY 321 ----- 321
Db 33993 AGTGGTGGGATGTTTCTCTCTAGCATATATGAAACAGTCTGTGTTCAGAGGACATGAT 33934
QY 322 -----ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleG1 337
|||:|||||
Db 33933 TTTTCTTCTAACAGTTTGGGAATAACAGCCAGCAATATCATTTATGGTGTCTGCAATGG 33874
QY 337 yAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleG1 357
|||:|||||
Db 33873 AAATGATGGCCACTTTAGCGAAGCTTAAATAATCCAGCTGACAGAGTGTATTAATAGG 33814
QY 357 yValGlyGlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTh 377
|||:|||||
Db 33813 TGTGGAGGATTTGACAATGATGATCATAGTCTCATTCATCTCGTGGCATGAGCAC 33754
QY 377 rTrpGluLeuProGly----- 382
|||:|||||
Db 33753 CTGGGAACCTCTCTCA-TGGGTACTGTATTTCTTGTGCTCTACATTCCTCTTTACAAA 33695
QY 382 ----- 382
Db 33694 CCCTACAATCGTAAAGATACAAAATAGCTTGTCTTCTTGTCAATTCCTAACTTCAAAGTTT 33635
QY 383 -----GlyTyrGlyArgMetLysProaspIleValThrTyrGlyAlaGlyValArgG1 400
|||:|||||
Db 33634 TGTTCACAGATGGCCGCTGTAACACCATGTTGTGTCATATGGCCGCCACATATGGG 33575
QY 400 ySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValVa 420
|||:|||||
Db 33574 GTCCAAGATCAGCAGCTGGTTGTTAAAGCCCTGTCTGGAACAAGTGTGCTAGTCTGTGT 33515
QY 420 lAlaGlyAlaValThrLeuLeuValSerThrVal-----GlnLysArgGluLeuVa 437
|||:|||||
Db 33514 TGTGTTATTTGTTTGGCTACTTCTTAAGTGTATTTCTCTGAAGCTAGAGGAAGATCTGCT 33455
QY 437 lAsnProAlaSerMetLysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValas 457
|||:|||||
Db 33454 CAATCAGCAGCAATGAGCAGCATTTGGTGAAGGTGCTGCTGCTTAAGCTTTCAGGCTCTAA 33395
QY 457 nMetPheGluGlnGlyHisGlyLysLeuAspLeu----- 468
|||:|||||
Db 33394 TATGTTATGACAGGCTCTGGAAGAGTGTGATCT-GTAAAGTACTTATCCATATTTGGTTAT 33336
QY 468 ----- 468
Db 33335 CACCATTGTACTCTTTATTTTATTTTAACTTCTAGTTCCTTTATTTTGTAGTCTCTAAT 33276
QY 469 -----LeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAl 482
|||:|||||
Db 33275 TTTTCCACTCTTTCTCAGATTACATCATATGAATCTTAAGAGCTACACCCCGGGC 33216
QY 482 aSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSe 502
|||:|||||
Db 33215 AAGCAATTTCCCGCAGCATTTCTTGATTAATATGATGTGCTCAATATCTCGGCCCTTTTGTG 33156
QY 502 rGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMe 522
|||:|||||
Db 33155 TCAGCCGCTATATGCGCGGCTATGCCCATTCATTTTCAACACGACAAATCTTAATGTTAT 33096
QY 522 tGlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnG1 542
|||:|||||
Db 33095 GGGTGTCTATGGCTATATTTGAAAGTCCACCAACATGGCACCT---GCAACAGGAGGAGG 33039
QY 542 yAspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGlyTyrLe 562
|||:|||||
Db 33038 AAATCTTCTGAGCATTCACCTTTAAGTACCGGATGTTATATGATGCGGTGGACTGTTATCT 32979
QY 562 uAlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyH1 582
|||:|||||

```
Db 32978 AGCTTACACATGAGATCAAGGAAGAGGTGCACAGTTCACAGGTGAATAGAGGCTAA 32919
Qy 582 sValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnTh 602
Db 32918 TGTAACTGTGAAGTCTATAGCCCA-----CCAGCCTCTGGAGAAAGTGGGCTAGAG 32865
Qy 602 rSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSerLysAr 622
Db 32864 GAGTACTGTCTCTCAGTTCGAACTGAAAGTGAAGTATCCCACTCCACACGACGGAAGC 32805
Qy 622 gValLeuTrpAspGlnTyrHisAsnLeuAlaGlyTyrProGlyTyrPheProArgAsnAs 642
Db 32804 TATTTTGGGATCAATTTACAGCAATTAATATCTCTCTGTTATATTCAGAGACTC 32745
Qy 642 nLeuArgMetLysAsnAspProLeuAspGlnPheValGlyPheHisPheArgPheAr 662
Db 32744 TTTGATCTCCGCNATGACATATCTGATGATGATGATGATGATGATGATGATGATGAT 32685
Qy 662 gAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLysLeuGlyAlaProPh 682
Db 32684 CATCAITGATACACATGTTACGTGATGCTGGGTACTATATCGAGACTCTTGGTTCGCCCT 32625
Qy 682 eThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyr 702
Db 32624 TACATGTTTTCATGCTCAGCAATATGGAACCTTTTGTGATGTTGACCTTGAAGATGATTA 32565
Qy 702 rPheProGluGluIleAlaLysLeuArgAspValAspAsn---GlyLeuSerLeuVa 721
Db 32564 CTTTCAGAGAATAATTGAATAACTCAGAGATGACGTTATCATATCTCGACTGGCTGTGT 32505
Qy 721 lIlePheSerAspTrpTyrAsnThrSerValMetAlaGlyLysValLysPheTyrAspGluAs 741
Db 32504 TGTGTTCTCAGTGTGATAATGTCGATACCATGCTGGATGAGATGATTCCTCGATGATAA 32445
Qy 741 nThrArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLe 761
Db 32444 CACACGAGTTGTGGACTCCAGTCACTGGAGTGCATATATCCACCACTTAATAACT 32385
Qy 761 uLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAl 781
Db 32384 TCTTCATCATTTGGGATGCTTTGGGGCAAAATCTGAATGGGATTTCAATGATTA 32325
Qy 781 aAsnHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGlyVa 801
Db 32324 CGTGAGCAGAGTCGATATGCTTCGGACAAACATGTCGGGTTCTCTCGCTGGTGGT 32265
Qy 801 lValIleThrGlnThrPheLysAsp----- 809
Db 32264 TTTGCACACTTTTCCTTTATATAGACAGCTCTGAGATGGCGCTACCCAAATCTACTGCT 32205
Qy 810 ----GlnGlyLeuGluValLysGlnGlnThrAlaValValGluAsnValPro----- 826
Db 32204 AACTGAGCATCAAAAGTAAATGAGAAACCAGTCTCTCCACTTCTAACTGTGCTTGGAT 32145
Qy 826 ----- 826
Db 32144 TTTCTTAATTAATTCAGTGAATTTGTTGTGATAATGTGGCTGTGTTTACAGGAAGACC 32085
Qy 827 ----lLeuGlyLeuTyrGlnIleProAlaGluGlyGlyValGlyValIleValLysTyrG 845
Db 32084 CTGCTGTTCTTGGGCTTTTGGAGATT-----CGTCAAGTTCGGTGGTGGTATATG 32034
Qy 845 lYasSerAsnCysLeuAspAspSerHisArgGlnLysAsnCysPheThrPleuLeuAspA 865
Db 32033 GAGACTCAAAATTTTGGATAGTACCCACATGTCACCACTGTCCACACTGTCTCTGAGGA 31974
Qy 865 lAlaLeuGlnTyrThrSerTyrGlyValThrProProSerLeu---SerHisSerGlyA 884
Db 31973 AATGCTAGATTTTACGATCTCAATATCAAGACCCCTGTACTTTTCTCAAAATTTGCTA 31914
Qy 884 snArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnH 904
Db 31913 AGAGATATTCACCAAGTAATCATAGACGAGAAGCAACATGCCATCCCGAAGCAATGATGA 31854
```

```
Qy 904 lSLeuHisArgTyrSerLysValLysValLeuGluAlaHisLeu 916
Db 31853 ATTTTTCACATACTCTCTCTGATCGGAAAGAGCTA 31816

RESULT 15
LOCUS AC017581 115884 bp DNA linear HTG 10-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
AC017581 pieces
ACCESSION AC017581
VERSION AC017581.1 GI:6554416
KEYWORDS HTG: HTGS_PHASE2
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
REFERENCE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
AUTHORS Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
TITLE Ephyroidae; Drosophilidae; Drosophila.
JOURNAL 1 (bases 1 to 115884)
COMMENT Adams,M. and Venter,J.C.
Direct Submission
Submitted (10-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
This sequence was identified as CDM:10211360 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
Location/Qualifiers
source 1..115884
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
BASE COUNT 33854 a 24803 c 24145 g 33082 t
ORIGIN
Alignment Scores:
Pred No.: 1,35e-102 Length: 115884
Score: 1772.00 Matches: 375
Percent Similarity: 56.61% Conservative: 122
Best Local Similarity: 42.71% Mismatches: 216
Query Match: 31.55% Indels: 167
Db: 2 Gaps: 14

US-09-830-837-6 (1-1052) x AC017581 (1-115884)
Qy 280 AsnAsnGlnValSerTyrThrSerTrpPheLeuAspAlaPheAsnTyrAlaIleLeuLys 299
Db 7 CATCGCAGGTTTCTTACACTTCTCGGTCTCGTATCGATCAATCACTACCGATATATAGG 66
Qy 300 LysIleAspValLeuAsnLeuSerIleGlyGlyProAspPheMetAspHisProPheVal 319
Db 67 AAAATAACATTTCTCAACCTTAGCATTTGGGGTCCCGACTTTATGGACTCGCCGTTGCTT 126
Qy 320 AspLysValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAsp 339
Db 127 GAAAGGTGTGGAACTGTGCGCTAATAATGTCATATATGATATCGCAGCAGGAATGAT 186
Qy 340 GlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGly 359
Db 187 GTCCTTGTACGGCAGCTTAACAATCTCTGGCATCGCAGCATGATGATGGCTGGT 246
Qy 360 GlyIleAspPheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrTrpGlu 379
Db 247 GCATTCAGTTGATGATTAATCCCAAGTTTATTCGAGAGGATGACACACGTGGAA 306
Qy 380 LeuProGlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArg 399
Db 307 CTTCCCTTAGCTACGAGCTATGGGACTCGATATTTGTCAGTCAGGAGCAAGTGGAA 366
Qy 400 GlySerGlyValLysGlyCysArgAlaLeuSerGlyThrSerValAlaSerProVal 419
Db 367 GGCAGTGTATGCGCAAGGGGTGCACAGCACTCTCTGGAACATCCGCTGCTCTCCAGTT 426
```

QY	420	ValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGluLeuValAsn	438	QY	744	nTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGlnLeuLeuSerVa	764
DB	421		439	DB	1495	ATGGTGACACCGACACCTGGTGGCGCAATATATCCAGCCTTGAATGATTTATTTGAAGCC	1554
QY	422		440	QY	765		784
DB	423		441	DB	1555	ATTGGGAATTCCTTTGGCGATTTGTGGTGAGGACATTTCAAACCTGGGCGACCATTC	1614
QY	424		442	QY	785		802
DB	425		443	DB	1615	AATGTACTATGCTAGTGGAGCCCAATCTGTAAGTTTCCCAATCAATCCAGGACGATATAT	1674
QY	426		444	QY	803		814
DB	427		445	DB	1675	AGTGGGCACAAAACCTGAATGACCAAGGACTTTCCGTAATTAATAATATCCTTTTGTAGT	1734
QY	428		446	QY	815		821
DB	429		447	DB	1735	TTCCACAGAAATACAACTTCCCTTTTCCAGATTATTAATTTCTAAACACCC-AGCAAGGT	1793
QY	430		448	QY	822	lValGlu---AsnValProIleLeuGlyLeuTyrGlnIleProAla-	835
DB	431		449	DB	1794	AGCAAACTAGATGCTACCTATTTTGGTATGTTCCAAACCAAGCGCAACAGTATTCAAAG	1853
QY	432		450	QY	835		835
DB	433		451	DB	1854	CAACGAGGAATCGTGGTCAATCGGAAAGCAATTTTGGCAGAGCTATATCCACAGATTA	1913
QY	434		452	QY	835		835
DB	435		453	DB	1914	CTCCACATTTAAGAACCGGTTTGTCTACTGCGAACGACGACGAAGATATCATGTTTGC	1973
QY	436		454	QY	836		848
DB	437		455	DB	1974	GAAGAACAATCATGAACACTAAGAATGAAGCAGCTATTTCCGCTATATGGGACTCCAA	2033
QY	438		456	QY	848	nCysLeuAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuG	868
DB	439		457	DB	2034	CTGCCCTGACCTCCACGATCTGGAGAGGCTTGTCTACTGCTGCTAATAACGCTTTTAGA	2093
QY	440		458	QY	868	nTyrThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgGlnArgPr	888
DB	441		459	DB	2094	TTTGGCAATTAAC-TCGCAC-	2112
QY	442		460	QY	888	oProSerGlyAlaGlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTy	908
DB	443		461	DB	2113		2134
QY	444		462	QY	908	rSerLysValLeuGluAlaHisLeuGlyAspProLysProArgProLeuProAla-	926
DB	445		463	DB	2135	AAATCGTATAACTGAATTTCAAAATTTAGAGAGAGCACCATTACCCCTTAGGATATCGCA	2194
QY	446		464	QY	927		935
DB	447		465	DB	2195	AGGTATTAATAATCTGTTTCACAGGACATAATTTGTGAACAATTTAAGTGGCTTGCACC	2254
QY	448		466	QY	935	oGlnProLeuAsnGluThr-AlaProSerAsnLeu-	948
DB	449		467	DB	2255	GACGAAGCAAAATACCCCGGAGGAAGAAATCTTCTATATAAGACGTACCACTACTGGA	2314
QY	450		468	QY	948	yHisGlnLysLeuLeuSerIleAspLysValValLeuProAsnPheArgSerA	968
DB	451		469	DB	2315	AAATGAAGGTATTTATATACATTTAGTTCAACTACAAAGCTCATATTTGTTGTTTCAG	2374
QY	452		470	QY	968	snArgPro-----GlnValArgProLeuSerProGlyGluSerGlyAlaTrpAspIleP	986
DB	453		471	DB	2375	AACACGAGATAAATTTAATCAAAATTTTATTTGGTGAGGAGATCGCAAACTAGGCAAA	2434
QY	454		472	QY	986	roGlyGlyIleMetProGlyArgTyrAsnGlnGluValGlyGlnThrIleProValPheA	1006
DB	455		473	DB	2435	ACAATGATATTTAACAGGANTCAATCCCGGTAGTCTAATGACTCCCAATATATCT	2494
QY	456		474	QY	1006	laPheLeuGlyAlaMetValValLeuAlaPhePheValValGlnIleAsn	1022

SGRGADEVHELKIKIDGYATVPSPVFMAMDVQVKEVKVFHEKRMNFFVRYRPAIVAAS
SVVAASVAIVATKKRG"

CDS
complement(join(33024..33137,33976..34006,34142..34194,
34276..34384,34509..34633,35297..35386,35465..35527,
35638..35729,35826..35907,36098..36271,36368..36449,
37285..37409,37497..37574,38232..38587,39496..39793))
/note="ESTs AU075322(C11109),D22430(C11109) correspond to
a region of the predicted gene.

Rice gene for soluble starch synthase (SSS1), complete cds
(exon1-15).(D38221)"

/codon_start=1
/protein_id="BAA81848.1"
/db_xref="GI:5295947"
/translation="MATAAGMGIGACILVAPQVRGRLBLORVRCVVAELSDGGS
AORLEAPLVKQVPLTELVPTSTPAQTSAPAPTPPIPDSCVGEIPELDLEGLI
EDSDITIEVASEQSEIMDKVQAOAKVTSVWFVGEASPKAKSGGIGDVCSLPI
ALALGRHVWVMPRYMGNALNKNFANATYKHKIKPCFGEHEVTFHEIRUSVDM
VFDVSHRPONLYGDNFGNQNFRITLLCYAACAEPLILLEGVYIYGOKCMFV
NMHSLVPIVLLAAKTRPGYIRDAKSVLVHNLAHQGVPEASTYDGLPPENYAL
EWPEWARHALDKGEAVNFKGAVVADRIYVTSQGYSWETTTAEGGQGLNELSS
RKSVLNGIVGIDINDMNPSTDKPLPHYSVDLSGRKAKBELQKELGLPRDPLV
IGFTLDYQKIGIDILKLAIPDLMDNIQFVMLSGDPPFEGWMRSTESGYRDKFRGM
VGFSPVSHRITAGCDIILMPSPFPCPNQILYAMQYGVVPHGGTGLRDTVENFP
PAKGGGQGWAFSPLETKIADGNFDIQGTQVLLGSNEARHVKRLYMGPCRLTV"
join(45936..46041,46221..46516,46748..46965,47016..47577,
48123..48381,50137..50150)
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA81849.1"
/db_xref="GI:5295948"

CDS
complement(join(50758..51068,53097..53115))
/note="hypothetical protein"
/codon_start=1
/protein_id="BAA81850.1"
/db_xref="GI:5295949"

Alignment Scores:

Pred. No.: 1 56e-75 Length: 170371
Score: 1355.00 Matches: 280
Percent Similarity: 45.97% Conservative: 91
Best Local Similarity: 34.70% Mismatches: 129
Query Match: 24.12% Indels: 308
Db: 8 Gaps: 8

US-09-830-837-6 (1-1052) x AB026295 (1-170371)

Qy 322 ValTrpGluLeuThrAlaAsnAsnValIleMetValSerAlaIleGlyAsnAspGlyPro 341
|||||
Db 166538 GTTTGGAGCTCACAAATAACATTATTATGCTATCAGCTATTGGAATGATGCGCT 166597
Qy 342 LeuTrpGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIle 361
|||||
Db 166598 CTTTATGACGCTAAATAACCGCTGATCAGAGTCAGCTCATTTGCTGTGGTATT 166657
Qy 362 AspPheGluAsnIleAlaArgPheSerArgGlyMetThrThrTrpGluLeuPro 381
|||||
Db 166658 GATTACATAACATATAGCTTCATTTCTCTCAGGAGCATGACTACTGGGAGCTTCT 166717
Qy 381
Db 166718 CATGGGTACATTTCTCACTTACCAGTACATATTTTCTTAGTCTGTCAGTGTGT 166777
Qy 381

Db 166778 TTTTTCGATATCTGGATCTCTAGCTTGAAGAAGCTAAAGCTTTGTATCATCTTCATTA 166837
Qy 381
Db 166838 GAACAGAGCTGTACAGTATATAAAACCTCTAAGCCCTAGTCTCAGTGGTTTTGTATATATAT 166897
Qy 381
Db 166898 CTTAAAAAAATTTGATGGAGTTTGGAGTACTTACATTCGACTTATCCCTATATATATATGT 166957
Qy 381
Db 166958 TGTAAAGGTCATTTGCTAGTGTGCTTACCAAACTTGCACCTGCAAGTCACACAAAA 167017
Qy 381
Db 167018 GATGTCATATGCTAATCTCTTTTAATCATACTAGCAGGTCAAGTATAGATTTAGGCTTT 167077
Qy 381
Db 167078 ACTGATGTGGTATCATATAAAGCTTTTTTACAAAGATTTCTCTCGAGAACTTTTGTAAA 167137
Qy 381
Db 167138 TCATATTCGATAATTTTATTTAGTTGAAATCAGATTTGTTGTGATATTAATTTATGG 167197
Qy 381
Db 167198 CAAGAAACTATTTCTGCCCTTGAACCTTAGGCATTTAATTAATATCTAGTAAAAAC 167257
Qy 381
Db 167258 TTGAATGATGAGCAATTTTGGGGGGGGGGGGGGGAGCGATGATCATTTTT 167317
Qy 381
Db 167318 TTACAATGCCATAGTACCATCTCTCTGTGTAGCTTTCAGATCCTTTGAATGGCTTA 167377
Qy 382
Db 167378 CTCATAAGTGTGGTTTCATTAATATCCAGG-TATGGCGGTGTAAACCTGATGTTGTGGC 167436
Qy 394 TyrGlyAlaGlyValArgGlySerGlyVallyGlyCysArgAlaLeuSerGlyThr 413
|||||
Db 167437 TATAGTCGTATTAATGTTCAAGATCAACAGTACACAGGTGTGTAACCCCTTTAGGCACC 167496
Qy 414 SerValAlaSerProValValAlaGlyAlaValThrLeuLeuValSerThrVal- 431
|||||
Db 167497 AGTGTGCAAGCCCGGTGTGCTGTGTGTAGTATGCTTGTGTGTATATACCAGAG 167556
Qy 432 --GlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuAlaSerAla 450
|||||
Db 167557 GAGCATCGAAATCCATCCTTAACCTGCAACAATGAAACAGCCCTTTGTAGAGGGTGT 167616
Qy 451 ArgArgLeuProGlyValAsnMetPheGluGlnGlyHisGlyLysLeuAspLeu- 468
|||||
Db 167617 TCTAGGCTTTTCAGGACCAACATGATGATGAGCAGGCGCTGGCAAGATTGACCTCTAAGTT 167676
Qy 468
Db 167677 TCTTGTCCCTTTTAAATGATCTTGTGGTATCCCTATGATAGTTAGTATTCATACTTC 167736
Qy 468
Db 167737 TTGTATCCCATCTTTCACACAGCAGGTTTGTCTTCAAAATTTAGAGCAGTATATC 167796
Qy 468
Db 167797 CTGTCAAGTTGGTTTGCCTTCTTTCATTTGTTGTGCTCAACAGATACTATTATTGTT 167856
Qy 469
Db 167857 TTGCTTTTCCAGCTGGCAATCATACGAAATCTTGAAAGATTACCAACCGCGTCAAGC 167916

	RESULT 18	AP002536	105815 bp DNA linear	PLN 21-MAR-2002
QY	484	LeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpPProTyrCysSerGln	503	
Db	167917	ATCTTCCTCAACCTTGACTACTCATGTGCCTATTCTGGCTTTTGGCGCAA	167976	
QY	504	ProIleTyrGlyGlyMetProThrValValAsnValThrIleLeuAsnGlyMetGly	523	
Db	167977	CCTCTGATGCTGGAGCTATGCCAGTGTCTCAATCTGACAAATCTGAACGGATGGG	168036	
QY	524	ValThrGlyArgIleValAspLysProAspTrpGlnProTyrLeuProGlnAsnGlyasp	543	
Db	168037	GTAATGGTATGTTAAGGATCCTCTGTATGGCAGCC---TCTGAGGATGTGGCAAT	168093	
QY	544	AsnIleGluValAlaPheSerTyrSerValLeuTrpProThrSerGlyTyrLeuAla	563	
Db	168094	ATCTTAGTGTTCACCTCACCTACCTACGATCATATGGCTTGCACAGGTATCTTGC	168153	
QY	564	IleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal	583	
Db	168154	CTGCACCTFACAAGTAAAGATGAAGGATCTCAGTTTTCAGGCATAATCAGTGGCAAGTC	168213	
QY	584	MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluInThrSer	603	
Db	168214	ACATTTGTCATTTACAGCCAGCA-----GCCATGGGGAAGCAGTCCACGAAGTAGT	168267	
QY	604	ThrValLysProIleLysValLysIleIleProThrProProArgSerLysArgVal	623	
Db	168268	TGATGTGTTCTTACCTCAAGGTCAAGTGTGGTGCACACCTGTAGTCCAGAAGAATA	168327	
QY	624	LeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAsnLeu	643	
Db	168328	TTGTGGACCATTTCATAATATCAATACCACTGCTGTTTTGTTCACAGAGATCACTT	168387	
QY	644	ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgasp	663	
Db	168388	AATGTGCATAATGATATCTTCTGCTGCGATGCGTACCTGTCACACGAATTTTCATATA	168447	
QY	664	MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr	683	
Db	168448	CTGTCAACATCTCTGAGAGATCGAGATACATACATGAGACCTTGGATCTCCACTTACC	168507	
QY	684	CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPhe	703	
Db	168508	TGCTTTGATGCTAGCAATATATGGGACACTGCTTATGTTGCTGACCTGAGGATGACTTC	168567	
QY	704	ProGluGluIleAlaLysLeuArgargasp---ValaspsnGlyLeuSerLeuValIle	722	
Db	168568	ACTGGAGGATTCAGAAAATTAAGATGATGTTGTACACAAAGGGCTTGGCGTTGTGTT	168627	
QY	723	PheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThr	742	
Db	168628	TTTGCAGAGTGTATCATGTCCACAAATGTTAAGATGACATTTCTTGTATGAAAAATACA	168687	
QY	743	ArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeu	762	
Db	168688	CGCAGTTGGTGACTCCAATTACTTGGGGTGCAAAATGTTCTCTGACCTTATGACCTTTG	168747	
QY	763	SerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsn	782	
Db	168748	GCACCATTTGGCATGCTTTGGGGACAAAGTACTAAGTGTGACTTCTCAATCAATGGT	168807	
QY	783	HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVal	802	
Db	168808	GAGCAGACCACTATGCTCTGGGACTGATATAGTGAATTTCCACAGGAGGTTTTTGG	168867	
QY	803	IleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGlu-----Thr	819	
Db	168868	-----CATAGCTTCCAACTCCAGGACAACTCTAAATTTCTCAGGATAACTCAAGAGC	168921	
QY	820	AlaValValGluAsnValPro	826	
Db	168922	GCAGATACACAAATATACCCCA	168942	

QY	468	-----	468
Db	945	TTGTCTTCAAAATTTAGAGGAGATATATCTGTCAAGTGGTTGGCCCTTCTTCTATT	1004
QY	469	-----	473
Db	1005	GTGTCTCAACAGACTCAATTTATTGTTTCTTTTCCAGCTGGCAATCATACGAA	1064
QY	474	IleLeuAsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGlu	493
Db	1065	ATCTTGAAGTTTACCAACCGCTGCAAGCATCTTTCCTTAATACCTTGACTTCACATGAT	1124
QY	494	CysProTyrMetIrrProTyrCysSerGlnProIleTyrGlyMetProThrVal	513
Db	1125	TGTCCCTATTTCTGGCTTTTGGCGCAACCTCTGTATGCTGGAGTATGCCAGTGGTC	1184
QY	514	ValAsnValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAsp	533
Db	1185	TTCATGCTACATCTGAACGGGATGGGTGATTGGTTATGTTAAGGATCTCTCTGTA	1244
QY	534	TrpGlnProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSer	553
Db	1245	TGGCAGCCC---TCTGAGGATGTCGGCAATATCTTAGTGTTCACCTACTACTCAGAT	1301
QY	554	ValLeuProProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAla	573
Db	1302	GTATATGCTTGGCAGGTATCTTCCCTGACCTTACATTAAGTAAAGATGAGGATCT	1361
QY	574	SerTrpGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThr	593
Db	1362	CAGTTTTCAGGCATAATCAGTGGCAAGTGCACATGTCTCAATTTACAGCCAGCA	1415
QY	594	GluSerLysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIle	613
Db	1416	GCCATGCGGGAAGCAGTCCACGAAGTAGTTCATGTCTTACCTGGAAGGTCAGGTG	1475
QY	614	IleProThrProProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyr	633
Db	1476	GTCCCAACACCTGTAGTCCAGAGAATATTGTGGGACCAATTTTCATATATCAATATAC	1535
QY	634	ProProGlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsn	653
Db	1536	CCATCTGCTTTTGTCCAGAGATTCATTAATGTGCATATGATATCTTCTGCTGGCAT	1595
QY	654	GlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyr	673
Db	1596	GCGCATCACTTGCACAGATTTTCATATATCTGTCAACATCTCAGAGATGCGAGATAC	1655
QY	674	PheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeu	693
Db	1656	TACATTGAGACCTTGGATCTCCATCTACCTTGTGCTTGTGCTGCTGCTGCTGCTGCTG	1715
QY	694	LeuMetValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgArgAsp	713
Db	1716	CTTATGCTTACCTTGGAGTACCTCAGTACGAGGATTCAGAGATTCAGAACTTAAGGATGAT	1775
QY	714	---ValAspAsnGlyLeuSerLeuValIlePheSerAspTyrTrpAsnThrSerValMet	732
Db	1776	GTGTGTACAAAGGGCTGGCTGTCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1835
QY	733	ArgLysValLysPheTyrAspGluAsnThrArgGlnTyrTrpMetProAspThrGlyGly	752
Db	1836	GTTAAGTACATCTTTTGTGAAATACACGAGTGTGGTGGCTTCAATTAAGTGGGGT	1895
QY	753	AlaAsnIleProAlaLeuAsnGluLeuLeuSerValTyrAsnMetGlyPheSerAspGly	772
Db	1896	GCAATGTTCTGCACTTAATGAACCTTTTGGCACCATTGTCATCTTGTGGGAGCAAA	1955
QY	773	LeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSer	792
Db	1956	GTACTAAGTGTGACTTCTCAATCAATGCTGAGCAGACCCACTATGCTTCTGGAGTAT	2015
QY	793	IleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeu	812

Db	2016	ATAGTGCANTTCCACAGAGGTTTGTG-----CATAGTCTCCACTCCAGACAAC	2069
QY	813	GluValLysGlnGlu-----ThrAlaValValGluAsnValPro	826
Db	2070	TCTAAATTTCTCAGGATTAACCTCAAGAGCGCAGATACACAATATCCCCA	2120
RESULT 19			
AC010662		113320 bp DNA linear HTG 18-FEB-2000	
LOCUS		Drosophila melanogaster clone RPC198-2K23, *** SEQUENCING IN	
DEFINITION		PROGRESS ***, 66 unordered pieces.	
ACCESSION		AC010662.5 GI:6996787	
VERSION		HTG; HTGS_PHASE1	
KEYWORDS		Drosophila melanogaster	
SOURCE		Drosophila melanogaster	
ORGANISM		Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Eukaryota; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
REFERENCE		1. (bases 1 to 113320)	
AUTHORS		Muzny, D.M., Adams, C.C., Bailey, M., Barbara, J., Blankenburg, K., Bodoty, B., Bouch, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C., Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C., David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N., Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D., Forcum-Tansley, J., Frantz, P., Ganes, R., Gorrell, J.H., Gorrell, L.L., Guvarra, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M., Hollway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M., Kelly, S., Kondejowski, N., Kong, Y., Kovar, C., Leal, B., Li, Z., Luchter, R., Liu, W., Liu, W., Logan, O., McLeod, M.P., Mei, G., Morgan, M., Morris, S., Nash, S., Nelson, A., Nguyen, R., Nguyen, R., Nguyen, S., Oswald, G., Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M., Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, R., Shah, E., Shen, H., Simon, M., Sparks, A., Stamps, A., Sugang, R., Taber, P., Taylor, T., Vasquez, L., Vinson, R., Vo, O., Wahban, M., Watlington, S., Weinstein, G., Weinstein, I.R., Williamson, A., Worley, K., Wren, J., Wrensford, G., Yu, W., Zhou, X., Nelson, D., and Gibbs, R.	
TITLE		Direct Submission	
JOURNAL		Unpublished	
REFERENCE		2. (bases 1 to 113320)	
AUTHORS		Worley, K.C.	
TITLE		Direct Submission	
JOURNAL		Submitted (17-SEP-1999) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA	
COMMENT		On Feb 18, 2000 this sequence version replaced gi:6056120.	
		----- Genome Center	
		Center: Baylor College of Medicine	
		Center code: BCM	
		Web site: http://www.hgsc.bcm.tmc.edu/	
		Contact: hgsc-help@bcm.tmc.edu	
		----- Project Information	
		Center project name: DRG8-2K23	
		Center clone name: RPC198-2K23	
		----- Summary Statistics	
		Sequencing vector: ML3; L08821	
		Chemistry: Dye-terminator Big Dye; 41% of reads	
		Assembly program: Phrap; version 0.980611	
		Consensus quality: 52892 bases at least Q40	
		Consensus quality: 72110 bases at least Q30	
		Consensus quality: 84532 bases at least Q20	
		Estimated insert size: 100525; sum-of-contigs estimation	
		Quality coverage: 1.2x in Q20 bases; sum-of-contigs estimation	

		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 66 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	
		* runs of N, but the exact sizes of the gaps are unknown.	

* This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 984: contig of 984 bp in length
 985 1004: gap of unknown length
 1005 2123: contig of 1119 bp in length
 2124 2143: gap of unknown length
 3254: contig of 1111 bp in length
 3255 3274: gap of unknown length
 3275 4526: contig of 1252 bp in length
 4527 4546: gap of unknown length
 4547 5431: contig of 885 bp in length
 5432 5431: gap of unknown length
 5432 5936: contig of 485 bp in length
 5937 5956: gap of unknown length
 5957 7052: contig of 1096 bp in length
 7053 7072: gap of unknown length
 7073 7889: contig of 817 bp in length
 7890 7909: gap of unknown length
 7910 9199: contig of 1290 bp in length
 9200 9219: gap of unknown length
 9220 10004: contig of 785 bp in length
 10025 10024: gap of unknown length
 10025 10938: contig of 974 bp in length
 10939 11018: gap of unknown length
 11019 11801: contig of 783 bp in length
 11802 11821: gap of unknown length
 11822 12546: contig of 725 bp in length
 12547 12566: gap of unknown length
 12567 13350: contig of 784 bp in length
 13351 13370: gap of unknown length
 13371 14218: contig of 848 bp in length
 14219 14238: gap of unknown length
 14239 15403: contig of 1165 bp in length
 15404 15423: gap of unknown length
 15424 16876: contig of 1453 bp in length
 16877 16896: gap of unknown length
 16897 18099: contig of 1203 bp in length
 18100 18119: gap of unknown length
 18120 19038: contig of 919 bp in length
 19039 19058: gap of unknown length
 19059 20658: contig of 1600 bp in length
 20659 20678: gap of unknown length
 20679 21926: contig of 1248 bp in length
 21927 21946: gap of unknown length
 21947 23095: contig of 1149 bp in length
 23096 23115: gap of unknown length
 23116 24605: contig of 1490 bp in length
 24606 24625: gap of unknown length
 24626 25352: contig of 727 bp in length
 25353 25372: gap of unknown length
 25373 27203: contig of 1831 bp in length
 27204 27223: gap of unknown length
 27224 28268: contig of 1045 bp in length
 28269 28288: gap of unknown length
 28289 29754: contig of 1466 bp in length
 29755 29774: gap of unknown length
 29775 31341: contig of 1567 bp in length
 31342 31361: gap of unknown length
 31362 32418: contig of 1057 bp in length
 32419 32438: gap of unknown length
 32439 33477: contig of 1019 bp in length
 33458 33477: gap of unknown length
 33478 34737: contig of 1260 bp in length
 34738 34757: gap of unknown length
 34758 35960: contig of 1203 bp in length
 35961 35980: gap of unknown length
 35981 37758: contig of 1778 bp in length
 37759 37778: gap of unknown length
 37779 38601: contig of 823 bp in length
 38602 38621: gap of unknown length
 38622 40035: contig of 1414 bp in length
 40036 40055: gap of unknown length

40056 41772: contig of 1717 bp in length
 41773 41792: gap of unknown length
 41793 43162: contig of 1370 bp in length
 43163 43182: gap of unknown length
 43183 44086: contig of 904 bp in length
 44087 44106: gap of unknown length
 44107 45469: contig of 1363 bp in length
 45470 45489: gap of unknown length
 45490 47159: contig of 1670 bp in length
 47160 47179: gap of unknown length
 47180 48716: contig of 1537 bp in length
 48717 48736: gap of unknown length
 48737 50174: contig of 1438 bp in length
 50175 50194: gap of unknown length
 50195 51756: contig of 1562 bp in length
 51757 51776: gap of unknown length
 51777 53606: contig of 1830 bp in length
 53607 53626: gap of unknown length
 53627 55470: contig of 1844 bp in length
 55471 56632: contig of 1142 bp in length
 56633 56652: gap of unknown length
 56653 59120: contig of 2468 bp in length
 59121 59140: gap of unknown length
 59141 60514: contig of 1374 bp in length
 60515 60534: gap of unknown length
 60535 61864: contig of 1330 bp in length
 61865 61884: gap of unknown length
 61885 64010: contig of 2126 bp in length
 64011 64030: gap of unknown length
 64031 65825: contig of 1795 bp in length
 65826 65845: gap of unknown length
 65846 67721: contig of 1876 bp in length
 67722 67741: gap of unknown length
 67742 70003: contig of 2262 bp in length
 70004 70023: gap of unknown length
 70024 72638: contig of 2615 bp in length
 72639 72658: gap of unknown length
 72659 75014: contig of 2356 bp in length
 75015 75034: gap of unknown length
 75035 77594: contig of 2560 bp in length
 77595 77614: gap of unknown length
 77615 80837: contig of 3223 bp in length
 80838 80857: gap of unknown length
 80858 83808: contig of 2951 bp in length
 83809 83828: gap of unknown length
 83829 85733: contig of 1905 bp in length
 85734 85754: gap of unknown length
 85754 88139: contig of 2386 bp in length
 88140 88159: gap of unknown length
 88160 91071: contig of 2911 bp in length
 91071 91090: gap of unknown length
 91091 95700: contig of 4610 bp in length
 95701 95720: gap of unknown length
 95721 99302: contig of 3582 bp in length
 99303 99322: gap of unknown length
 99323 103945: contig of 4623 bp in length
 103946 103965: gap of unknown length
 103966 108234: contig of 4269 bp in length
 108235 108254: gap of unknown length
 108255 113320: contig of 5066 bp in length.

FEATURES Location/Qualifiers
 source 1. .113320

Alignment Scores:

Pred. No.: 2.95e-43 Length: 113320

Score: 850.50 Matches: 173

Percent Similarity: 64.00% Conservatives: 67

Best Local Similarity: 46.13% Mismatches: 96

Query Match: 15.14% Indels: 44

DB: 2 Gaps: 4

US-09-830-837-6 (1-1052) x AC010662 (1-113320)

QY 782 nHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPhe 796
 Db 41145 ATATTATATATAGTGTAGGAGTATTAATGTTAAGTTT 41187

RESULT 20
 LOCUS AK054744
 DEFINITION Homo sapiens cDNA FLJ30182 fis, clone BRACE2001154, highly similar to intraluminar subtilisin-like proteinase 51P.
 ACCSSION AK054744
 VERSION GI:16549348
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens cerebellum cDNA to mRNA, clone lib:BRACE2
 clone:BRACE2001154.

ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE
 1 Nishi, T., Nakagawa, S., Senoh, A., Mizuguchi, H., Inagaki, H., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuo, Y., Nagai, K. and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 3333)
 Isogai, T., Otsuki, T. and Sugiyama, T.
 Direct Submission
 Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory, 1532-3 Yana Kisancho, Chiba 292-0812, Japan (E-mail: genomics@helix.co.jp, tel: 81-438-52-3975, fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES
 source
 Location/Qualifiers
 1..3333
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="BRACE2001154"
 /tissue_type="cerebellum"
 /clone_lib="BRACE2"
 /note="cloning vector: pME18SFL3"
 BASE COUNT 785 a 816 c 888 g 844 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6,66e-43 Length: 3333
 Score: 812.00 Matches: 194
 Percent Similarity: 31.65% Conservative: 0
 Best local Similarity: 31.65% Mismatches: 1
 Query Match: 14.46% Indels: 419
 DB: 9 Gaps: 1

US-09-830-837-6 (1-1052) x AK054744 (1-3333)

QY 858 AspCysPheTrpLeuLeuAspAlaLeuLeuGlnThrSerTyrGlyValThrPro 877
 Db 818 GACTGCTTTTGGCTTCGGATGCCCTCCAGTACATCATGATGGTGACCGCT 877

QY 878 SerLeuSerHisSerGlyAsnArgGlnArgProSerGlyAlaGlySerValThrPro 897
 Db 878 AGCCTCAGTCACTCTGGGAACCCAGCCGCCCTCCAGTGGAGGAGCTCACTCA 937

QY 898 GluArgMetGlu----- 901
 Db 938 GAGAGAGTGA-AGGTGAGTGGATGGCAGGATGCTTGGGGAGGAATGGATCCAGCT 996

QY 457 AsnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsn 476
 Db 40082 AACATTTTGGAGGAGGAGCTGGAAACTGAATTTGCTGAAGAGTATGCGATGCTG 40141

QY 477 SerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyr 496
 Db 40142 TCATACAAACCAAGATAACCTTATTCGCGCATACCTTGACTTCAACCCAA--AACATAT 40198

QY 497 MetTrpProTyrCysSerGlnProIleTyrTyrGlyMetProThrValValAsnVal 516
 Db 40199 ATGTGGCTTATAGCTCCCAACCTCTGCTACTGGAAGCTCGCTGCTATGCAACGTT 40258

QY 517 ThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGlnPro 536
 Db 40259 ACATATCTATGATGTCGTCTGTCACAGTCATATGATGGTCCCTCAATGGATTCCTC 40318

QY 537 TyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTrp 556
 Db 40319 GATTTTGGAAACCAAGTCACTTTCTTCAAGTATCTGACCAAGTTTCCCTATCGTTGG 40378

QY 557 ProTrpSerGlyTyrLeuAlaIleSerIle----- 566
 Db 40379 CCGTGGACCGGTGGATGTCAGTTTATTTGGTGAAGTATGAAATACCAACCTATATA 40438

QY 566 ----- 566

Db 40439 AAGGATATATGGGTACATATTGATATGAAAGTCTCAAGCCATCTATACCAACAATTT 40498

QY 567 -----Ser-ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVa 583
 Db 40499 TTGTTTACCACTGTATAAAGGAGGAGAACTTTGAAGTGTGTTGTAAGCAAGTAT 40558

QY 583 I MetIleThrValAlaSerProAlaIleThrGluSerLysAsnGlyAlaGluGlnThrSe 603
 Db 40559 CACCTAGTTTGGAAAGCTTTAAACAGCACCAACAACT-----CATGTTCAC 40609

QY 603 rThrValLysProIleLysValLysIleIleProThrProArgSerLysArgVa 623
 Db 40610 AGAAGTCGACTTTCCTTAAACAATAAGGTTTACTTCCAAACCCGCAAGAAAGAGAT 40669

QY 623 ILeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLe 643
 Db 40670 TTTTGGATCAGTACCACAGCTTAAGTATCCACCGCTATATCCACGAGATGATCT 40729

QY 643 uArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAs 663
 Db 40730 CAAGTTAATCTAGTCTCTGACTGGAGGAGACCAATATACACAACTTTAGGCA 40789

QY 663 pMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheTh 683
 Db 40790 CATGTATACATTTAGCAATGTTGGCTACTCATTTGATGTTTGGAGAACCTTCAC 40849

QY 683 rCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlnTyrPh 703
 Db 40850 CTGCTTTCAATCGCTCGGATTTAGCGCGCTTATGATTTGACCTGAGAGAGGTTTGG 40909

QY 703 eProGluGluIleAlaLysLeuArgAspVal---AspAsnGlyLeuSerLeuValI 722
 Db 40910 CGACGAGGAAATAAACGCTTTACAGGAAACGCTGTATAAAGAGGCTTGTATGCTGCT 40969

QY 722 ePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnTh 742
 Db 40970 ATTCGGAGACTGGTATAACACCACTGTGATGAAATAAATAATCTTTGACGAGAC-AC 41028

QY 742 rArgGlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLe 762
 Db 41029 CCACAACTGGTATATT-GATATGTTGGTGA-AATAATTA-GTATTGATGATTTATT 41085

QY 762 uSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAs 782
 Db 41086 GAAGGA-TTTGGAAATGCTTTTGGGATTTTGTGGTGGAGGATATTTTAATTTGGTCT 41144

QY	901	-----	901	QY	901	-----	901
Db	997	GAACATGATCTACTTCAGCCAGTGGCTGTGCCTGCCAGCAAGTGGGGCCAGTTGTGTG	1056	Db	2077	TAGACACAGGGGCTCTGTCTGTCTGTGGAAAGACAGACAGACTCATATAAGGCCCT	2136
QY	901	-----	901	QY	901	-----	901
Db	1057	CGCGCTGAGTGGTTTTTCGTAATGCTGCGGTAGCAGCAGCACTCAGCGGCTCTAACCTG	1116	Db	2137	CCTTTTCCCTTTCATTTAGTAATAGAGATGATTTTCCCTTTTGTGTGCTGTCTGCTA	2196
QY	901	-----	901	QY	902	-----	901
Db	1117	CCGAGAGGCCATCCTTAGTCCCGAGAGGCCATGATCGAGGACCCAGTGTGCAGCTGCTT	1176	Db	2197	ACCACAGGAACCACTTCATCGGTACTCCAAGGTTCTGGAGGCCCACTTTGGGAGACCCA	2256
QY	901	-----	901	QY	920	LysProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGlnProLeuAsn	939
Db	1177	TCCCACGTGCACCCGACCCTCAGCCACTCTGAGTGTGTTACCGTCCCGCTCTCAG	1236	Db	2257	AAACCTCGGGCTCTACACAGCTGTCCAGCTGTCTTGGGCCAAGCCACACGCCCTTTAAAC	2316
QY	901	-----	901	QY	940	GluThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAspLeuAspLys	959
Db	1237	GTGAGGACAGTGGCTCCTACAGATCTTTCAGGAAGATCACCAACATTTCTAGAGCACTGC	1296	Db	2317	GAGACGGCGCCAGTAACTTTGGAACATCAGAAGCTACTCTCCATTGACCTGGACAAG	2376
QY	901	-----	901	QY	960	ValValLeuProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGlu	979
Db	1297	CAGTCCCATGCAGCAGGCTGGGATGCTCCATGGATTACCTCACTGTCCCTCTCAGTA	1356	Db	2377	GTGGTGTATACCACTTTCGATCGAATCGCCCTCAAGTGAGGCCCTTGTCCCTGGAGAG	2436
QY	901	-----	901	QY	980	SerGlyAlaTrpAspIleProGlyGlyIleMetProGlyArgTyrAsnGlnGluValGly	999
Db	1357	ATCCTGTGAAATAGATGTTATTATTCACAGTTTCCAGATGAGGAATGAGAGCTGAGAG	1416	Db	2437	AGCGGCCCTGGGACATTCCTGGAGGATCATGCTGGCCGCTACACAGGAGGTGGGC	2496
QY	901	-----	901	QY	1000	GlnThrIleProValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValVal	1019
Db	1417	AAATTAGATAACTGCCCTGAGATCGCTGGCGGTATGTGGTGGAGCCTCAACCCGCC	1476	Db	2497	CAGACCATTCCTGCTTCTTGGCTTCTCTGGAGCCATGGTGTCTCTGGCTCTTTGTGTGTA	2556
QY	901	-----	901	QY	1020	GlnIleAsnLysAlaLysSerArgProLysArgArgLysProArgValLysArgProGln	1039
Db	1477	TCTGGGGCCTGTGCTGGTGGCGTGGGGCCCGCAGCTGCTTGGCACATTAGTGGGTGGTA	1536	Db	2557	CAATCAACAGGCCAAGACAGCCGAGCGAGGAGGCCAGGGGTGAAGCCGCCGAG	2616
QY	901	-----	901	QY	1040	LeuMetGlnGlnValHisProLysThrProSerVal	1052
Db	1537	TCAGTGGCTAAGTCCCTGGTGTGCTGGTCCCTTACAGCGCTGCCCGTAGAAGCTCATG	1596	Db	2617	CTCATGCAGCAGGTTACCCGCCCAAGACCCCTTCGGGTG	2655
QY	901	-----	901	RESULT 21			
Db	1597	GGAACACAGGCATCATTTAAATTTTCCAGAGCTGCGTTTAAAAAGTAAAAAGAAACA	1656	AC010056	129955 bp	DNA	linear
QY	901	-----	901	LOCUS			
Db	1657	CAGATGAACATATTTTAATAATATACTTTGAGCCCAAGACATGAAGAATATTAGAAATTAA	1716	DEFINITION	Drosophila melanogaster chromosome 3L/79A3 clone RPC198-25D3, ***		
QY	901	-----	901	ACCESSION	AC010056	GI:6056152	
Db	1717	ACAGGTGATACCTATTTCACTTGCTCAGTACGTACGTGTGGTGGCACAATGGTGCAGA	1776	VERSION	AC010056.5		
QY	901	-----	901	KEYWORDS	HTG; HTGS_PHASE1		
Db	1777	TCTAAGAAATAGAAAGAAATTTGAAATATGACAGAGTGTTTATTATTGGCCAGAGAATTA	1836	SOURCE	Drosophila melanogaster.		
QY	901	-----	901	ORGANISM	Drosophila melanogaster.		
Db	1837	AATATTTTAGAGGGAAGTATTTTTCATGAACCCGGTCTGTAGCTTAGATTTCGCAGG	1896	REFERENCE	1 (bases 1 to 129955)		
QY	901	-----	901	AUTHORS	Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,		
Db	1897	TGTAGCTTGGTGTCTCAGTGGTGTGCACCAAGTTTGACAGCCAGCTGCTGTGCACAC	1956		Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,		
QY	901	-----	901		Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,		
Db	1957	ATCTGCCCATCTTCAAGGGAACGGTCTGTGATTTTCTCACCCACCTTCAGCAATGGAACA	2016		Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,		
QY	901	-----	901		Forum-Tansey, J., Frantz, P., Ganesh, R., Gorrell, J.H., Gorrell, L.L.,		
Db	2017	CACATTGTCGGGAGGAACCAATCAGCGGTGGCTCCTGGGCTCATGGAGATAGTTAA	2076		Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,		
					Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,		
					Kelly, S., Kondejowski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,		
					Lichtharge, O., Liu, J., Liu, W., Logan, O., Lu, J., Lucier, R.,		
					Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M., Morris, S.,		
					Nash, S., Nelson, A., Nguyen, R., Nguyen, N., Nguyen, S., Oswal, G.,		
					Parish, B., Paxton, S., Payton, B., Perez, L., Pu, L.L., Quiles, M.,		
					Reiter, D., Rives, M., Samuel, S., Say, J., Scherer, S., Shah, E.,		
					Shen, H., Simon, M., Sparks, A., Stamps, A., Sugeng, R., Taber, P.,		
					Taylor, T., Vasquez, L., Vinson, R., Vo, Q., Wahbah, M., Watlington, S.,		
					Weinstock, G., Weinstock, I.R., Williamson, A., Worley, K., Wren, J.,		
					Wrensford, G., Yu, W., Zhou, X., Nelson, D. and Gibbs, R.		
					Direct Submission		
					TITLE		


```

QY 682 eThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlu 702
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18159 CACCTGCTCAATGCGCTGGAATATGGCGGCTTATGATGTTGACCCNTGAGAGGGGT 18218
*
QY 702 rPheProGluGluLeuAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValII 722
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18219 TTGGCGCAGCAGGACAAATAGCTGCTTACGATG--GACAATGCATACGACTTGCT 18276
*
QY 722 ePheSerAspTyrTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsn 742
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18277 ATTAGCAGTACGTTTCATCCCTGTTTATGTAATGTCATGGGAAGGAGCAAC 18336
*
QY 742 rArgGlnTyrTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeu 762
|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18337 AATAAGAGGCTTCCTGGCCAGGACGACAGATTTATGGCCGCCAATATGTCAAAATATT 18396
*
QY 762 euSerValTrpAsnMetGly 768
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 18397 AAGAGACATGGACAATGGC 18416
*
RESULT 22
AC011909
LOCUS Drosophila melanogaster chromosome 3L/79A3 clone RPC198-25D3, ***
DEFINITION SEQUENCING IN PROGRESS ***, 1 unordered piece.
ACCESSION AC011909.1 GI:6056127
VERSION HTG; HTGS_PHASEL.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 129955)
AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbara,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
Guevara,W., Harris,K., Hernandez,J., Hodgson,A., Hoques,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondesewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichtarge,O., Liu,J., Liu,W., Logan,O., Lu,J., Lucier,R.,
Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M., Morris,S.,
Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G.,
Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M.,
Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E.,
Shen,H., Simon,M., Sparks,A., Stamps,A., Sucgang,R., Taber,P.,
Taylor,T., Vasquez,L., Vinson,R., Vo,Q., Wabwah,M., Watlington,S.,
Weinstock,G., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrensford,G., Yu,W., Zhou,X., Nelson,D. and Gibbs,R.
Direct Submission
Unpublished
2 (bases 1 to 129955)
Worley,K.C.
Direct Submission
Submitted (16-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 820: contig of 820 bp in length
* 821 1892: contig of 1072 bp in length
* 1893 3138: contig of 1246 bp in length
* 3139 4261: contig of 1123 bp in length
*
FEATURES
source
1..129955
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L/79A3"
/clones="RPC198-25D3"
BASE COUNT 35760 a 28458 c 29103 g 36584 t 50 others
ORIGIN
Alignment Scores:
Pred. No.: 2.8e-38 Length: 129955
Score: 774.50 Matches: 163
Percent Similarity: 59.28% Conservative: 67
Best Local Similarity: 42.01% Mismatches: 118
Query Match: 13.79% Indels: 42
DB: 2 Gaps: 4
US-09-830-837-6 (1-1052) x AC011909 (1-129955)
QY 417 SerProValValAlaGlyAlaValThrLeuLeuValSer---ThrValGlnLysArgGlu 435
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 17272 TCCTCCAGTTGTTGCAGGGGCTGCTGCTATTAAGCGGTTCAGAAATCGAC 17331
*
QY 436 LeuValAsnProAlaSerMetLysGlnAlaLeuAlaSerAlaArgLeuProGly 455
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

Db 17332 TACATAAACCCAGCATCTCTTAAGCAGGTACTCATTTGAAGTGGCGAGAAATGCGCGCAT 17391
QY 456 ValasnMetPheGluGlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeu 475
Db 17392 TATAACATGTTTGGCAGGAGCTGGAAAGCAATTTGCTGAAGAGTATGCGAGTATG 17451
QY 476 AsnSerTyrLysProGlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysPro 495
Db 17452 CTGTCATACAAACCAAGATAACCTTATTCGGGCATACCTTGACTTCACCAA---AAC 17508
QY 496 TyrMetTrpProTyrCysSerGlnProIleTyrGlyMetProThrValValasn 515
Db 17509 TATAATGGGCTATAGTCCCAACCTCTCTACTATGGAACTCCGTCGCTATTCACAA 17568
QY 516 ValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAspTrpGln 535
Db 17569 GTTACCATCTCAATGTTATCTCTGTCACAGTCATATAGTTCGATCCCTTAATGGATT 17628
QY 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeu 555
Db 17629 CCCGATTTCGAAACCAAGGTCAGTTCTTCAAGTATCTGCACAAAGTTTCGCTATCGT 17688
QY 556 TrpProTrpSerGlyTyrLeuAlaIleSerIle----- 566
Db 17689 TGGCGGTGACCGGTTGGATGTCAGTTTATTTGTTGATGATGAAATAGCAACGCTAT 17748
QY 566 ----- 566
Db 17749 ATAAAGGATATATGGGTGATATGATGATAAGTGTCTAAGCCATCTTATACCAACAA 17808
QY 567 -----Ser-ValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHI 582
Db 17809 TTTTGTGTACAGCTGTAAAGGAAGGAGAAACTTTGAAGGTGTTTGAAGAGGAG 17868
QY 582 sValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluTh 602
Db 17869 TATCACCCCTAGTTTGTGAAGCTTTTAAACAGACACCACCAAGCAACT-----CATGT 17919
QY 602 rSerThrValLysLeuProIleLysValLysIleLeuProThrProArgSerLysAr 622
Db 17920 TACAGAGTCGACTTCTTTTAAACATAAAGGTACT-CCAAACCGCCAAAGAAACAGAG 17978
QY 622 gValLeuTrpAspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArgAspAs 642
Db 17979 GATTTTATGGATCGTACACAGCTTAAGGTATCCACGCGCTATATCCAGAGATGA 18038
QY 642 nLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheAr 662
Db 18039 TCTCAAAGTTAACTAGATCCTCTGGACTGGAGGCGAGACCATATACACAACTTTAG 18098
QY 662 gAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPh 682
Db 18099 GGACATGTATACATTTACGAAATGTTGGCTACTCATTTGATGTTTGGCAGAACCTTT 18158
QY 682 eThrCysPheAspAlaSerGlnTyrThrLeuLeuMetValAspSerGluGluGluTyr 702
Db 18159 CACCTGCTTCATCCCTCGGATATGCGCGTATTGATTGATTGATCCNTGAGAGAGGT 18218
QY 702 rPheProGluGlnIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValI 722
Db 18219 TTGGCGCAGGAGCAATAAGCTGCTTACGATG-GACAATGATGACGACTTGTCT 18276
QY 722 ePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnTh 742
Db 18277 ATTAGCCAGTACGTTTCATCCCTGTTTATGTAATATGTCATGGGAGGAGGAGCAACAC 18336
QY 742 rArgGlnTrpTrp-MetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeu 762
Db 18337 AATAAGAGGCTTCTCGGCCAGGACCAAGCATTTATGGCGCGCCCAATATATGTCANAAATATT 18396
QY 762 euSerValTrpAsnMetGly 768
Db 18397 AAGAGACATGGCAATGGGC 18416

```

```

RESULT 23
AC040169/c
LOCUS
DEFINITION
SEQUENCE, 4 ordered pieces.
AC040169
AC040169.6 GI:18139274
HTG: HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 214287)
Sequencing of Human Chromosome 16
DOE Joint Genome Institute.
2 (bases 1 to 214287)
Unpublished
DOE Joint Genome Institute.
Direct Submission
Submitted (11-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jan 12, 2002 this sequence version replaced gi:13786410.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 9563: contig of 9563 bp in length
* 9564 9663: gap of unknown length
* 9664 26317: contig of 16654 bp in length
* 26318 42789: gap of unknown length
* 42790 42889: contig of 16372 bp in length
* 42890 214287: gap of unknown length
* 214287 214287: contig of 171398 bp in length.
* Location/Qualifiers
* 1. 214287
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="16"
* /clone="RP11-505K9"
BASE COUNT 57135 a 53978 c 52644 g 50229 t 301 others
ORIGIN
Alignment Scores: 5,55e-35 Length: 214287
Pred. No.: 728,00 Matches: 269
Score: 18,548 Conservative: 8
Percent Similarity: 18.01% Mismatches: 19
Best Local Similarity:

```


Db	127203	GGGGTATTGCTAAGCTCTCTGTGACACACAGATGGTGGCAGCAATTTACTTGAAG	127144	Db	126124	CATCTAGTGAAGCTGACACACACTCTCTGTATGAGAAAGTGAAGCTTTTATGATGAAA	126065
QY	652	-----	652	QY	741	nThr-----	742
Db	127143	CGACTGCTACAGTGAATGCTCCAGGACACATTAACATGCTTTTGCCATTTCTCTGGTC	127084	Db	126064	CACAAGGTACTGTTGATACGCATTTGATTTGGATAAGACTTTTGGGAATAAAATCTTA	126005
QY	652	-----	652	QY	742	-----	742
Db	127083	ATTTTGTCTTTTCATGCACCTGCCAGGCACATGGTTTCTTCTTATCAGCCATTTCCGTG	127024	Db	126004	CCCAGCAATACTTTAAGTGAATTTGACCTGTTGGTGGTCTTTCCAGTTGGCACAGACA	125945
QY	652	-----	652	QY	742	-----	742
Db	127023	TGTCAGACGACAGACACAGGTCCTTAAGCTGATGCCTGTGCCAGTGGCCAGCCATC	126964	Db	125944	CATTGAGAGCTTTGATTAGAAAAAAGCCCTTTTCGACTGAGTGGCTGCTACGCCCTG	125885
QY	652	-----	652	QY	742	-----	742
Db	126963	CAGGCTTCTCTGTGGCCATGGCATGGTCTCTGCAGAGCCGCCCTGTGGCCAGATA	126904	Db	125884	TAATCCAGCACTTTGAAGGCTGAGTTGGGCAGATCGGTTGAACCTGGGAGGCCAAGGC	125825
QY	653	-----As 653	653	QY	742	-----	742
Db	126903	TGCCTTTATGCCAGTAAGATCGATCATTTTAATGCTGTGGCCTATTGTTTTCAGGAA	126844	Db	125824	TGCAGTGAAGTGAAGCCACTGTACTCCAGCTGGGTGTCAGAGAGACCCTATCTC	125765
QY	653	nglyAspHisIleHisThrAspMetTyrGlnHisLeuArgSerMetGlyTy	673	QY	742	-----	742
Db	126843	TGTTGATCATATCCACCAATTTTCCAGGATATGTACCAAGCATCTGGAAGCATGGGCTTA	126784	Db	125764	AAAGAGAAAAAATAAACACACAAAAAAGAAATAGTCTTTATACAGTAGAGAGGTTCTTG	125705
QY	673	rPheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThr--	692	QY	742	-----	742
Db	126783	CTTTGTAGAGTCTCTCGGGCCCCCTTCACGTGTTTTGATGCCAGTCAGTATGGTAA--GT	126725	Db	125704	AAAGTTGCTCATAGAGTAAGATGTTTATGTAGATATGAGTACTCTTAATAATATAGAGGT	125645
QY	692	-----	692	QY	742	-----	742
Db	126724	GGCAGTGTGTGTAGCCAGATGGTCTTTTCGGCCACTTTGGATGTTTGTAAATGTCG	126665	Db	125644	TCATGTCTTCTTCTTAACTTAGATATGAGTACTTCAGGTTTAAACTTCTGTTTAA	125585
QY	692	-----	692	QY	742	-----	742
Db	126664	AGAGTGAATCCAGTTTTTCAGTGATAGCTACCAAGGAGTATGCGCGTGGCCCCACA	126605	Db	125584	TTCTTTTCCACTCTCTGTTACTGAGTATATAATAATTAATGCACAAAGAACTGTTT	125525
QY	692	-----	692	QY	742	-----	742
Db	126604	GCTGGGTGAGCCTGGGCGAGGCTCTGAACAGGTGGTGGCGGTGACATTCAGCTGAA	126545	Db	125524	CITTAAGAATATTTGAGTAAATTTATGTATATTTAAATATTCATATTTACAATAATTA	125465
QY	692	-----	692	QY	742	-----	742
Db	126544	ATACAGACAGGCTGATTTGGAGACAGGACATTCGTACTCTTTTCTTTAAGTTAAAG	126485	Db	125464	ATTTACTTATATTTATAGTAAATTAATTTACTCAAGGGAATATGTCCTTCCCTTATTTG	125405
QY	692	-----	692	QY	742	-----	742
Db	126484	CAGCTCTGTATATTTTGTCTCTCTCAAGTGAATGAAATCTTTATCTGAGCACCA	126425	Db	125404	GGATCAAGGACTCCCAATGTATATAACATATACACCTTTATGTTTGAATCGGTTTCT	125345
QY	692	-----	692	QY	742	-----	742
Db	126424	CAAGTGAATCAGATCAGTTAGTCTGTGTATTGCTGGATCTGCTCTACTTTT	126365	Db	125344	TTAGAATCTGTGACTGAACATTTCTATGCTTTTCTCATCTCTCTCTCTCTCTCTCTCT	125285
QY	692	-----	692	QY	742	-----	742
Db	126364	TAGTCCCTTTTAAAGTGTACAGGGTGGCACAACACAGACTGTTAGTGTACGTAACCTG	126305	Db	125284	CTTGCTGGTCTAGGCTAAGCAGTATCTCTGGTGCATTTTCGGTAAAGTACTGATTTTCA	125225
QY	692	-----	692	QY	742	-----	742
Db	126304	AGACATGATCTGTGACACTGTTTCTGCGGTGGTGTGCTGCTGCTGCTGCTGCTGCTG	126245	Db	125224	TTTCATAAATAAATTTTGTCTTTAATTAATTTACATACATCAGCCGAGCACTTAAATTTG	125165
QY	693	-----LeuLeuMetValAspSerGluGluGlu 701	701	QY	742	-----	742
Db	126244	GTTCACGCTTTGTTTGTTCATCAGGCATTTGCTGATGGTGGACAGTGGAGGGA	126185	Db	125164	CAGTTTACCCAAACCCACAGACTGGTTTCAGTTGAAGCCTTAGTGTGGAGGACTGGA	125105
QY	701	uTyrPheProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuVa	721	QY	742	-----	742
Db	126184	GTACTTCCCTGAAGATCCCAAGCTCCGAGGAGCGTGGACACAGCCCTCTCGCTCGT	126125	Db	125104	TTGTGGCAGGAGAAATGAGTTTGTACACAGACTAAATACTGTGTATGGAGTGCTTT	125045
QY	721	lIlePheSerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAs	741	QY	743	-----ArgGlnTrpTrpMetProAspTh 750	750
				Db	125044	CATTCCCTGTTCCCTGTGTGCTGTTTACCACCTTTCCCGAGCGAGTGTGTGATCCGGATAC	124985


```

QY 750 rGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSe 770
|||||
Db 124984 CGGAGGAGCTACATCCAGCTCTGAATGAGCTGCTGCTGTGTGGAACATGGGTTTCAG 124925
|||||
QY 770 rAspGlyLeuTrpGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGI 790
|||||
Db 124924 CGATGCCCTGTATGAAGGGGAGTTCCACCTGGCCACCATGAC----- 124882
|||||
QY 790 yCysSerIleAlaLysPheProGluAspGlyValValIle 803
|||||
Db 124881 ----ACTAAGGCTCTGTTTCTCTGAG---GGCGTGGTCTTA 124849
|||||
RESULT 24
AB047598
LOCUS AB047598 2905 bp mRNA linear PRI 01-SEP-2000
DEFINITION Macaca fascicularis brain cDNA, clone:QnpA-10050.
ACCESSION AB047598
VERSION AB047598.1 GI:9929930
KEYWORDS fis (full insert sequence).
SOURCE Macaca fascicularis adult male brain parietal lobe cDNA to mRNA,
clone_lib:macaque brain cDNA library QnpA clone:QnpA-10050.
ORGANISM Macaca fascicularis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
Cercopithecinae; Macaca.
REFERENCE 1 (sites)
Osada, N., Hida, M., Kusuda, J., Tanuma, R., Iseki, K., Hirai, M.,
Terao, K., Suzuki, Y., Sugano, S. and Hashimoto, K.
Isolation of full-length cDNA clones from macaque brain cDNA
libraries
Unpublished
JOURNAL 2 (bases 1 to 2905)
Hashimoto, K., Osada, N., Hida, M., Kusuda, J. and Sugano, S.
Direct Submission
Submitted (23-AUG-2000) Katsuyuki Hashimoto, National Institute of
Infectious Diseases, Division of Genetic Resources; 23-1, Toyama
1-chome, Shinjuku-ku, Tokyo 162-8640, Japan
(E-mail: khashi@nih.go.jp, URL: http://www.nih.go.jp/yoken/genebank/,
Tel: 81-3-5285-1111 (ex. 2120), Fax: 81-3-5285-1181)
Lab host: TOP10
Vector: PME18S-FL3 (Acc.No. AB009864)
R. Site1: DraIII (CACTGTGTG)
R. Site2: DraIII (CACCATGTG)
Description: 1st strand cDNA was primed with an oligo(dT) primer
[ATCTGGCCTTTTCTTTTCTTTT]; double-stranded cDNA was synthesized
using specific 5' and 3' primers and amplified by PCR. The PCR
product was digested with SfiI and size selection was performed to
exclude fragments <1.5kb. The SfiI-digested PCR product was cloned
into distinct DraIII sites of PME18S-FL3. XhoI sites just outside
the DraIII sites can be used to isolate the cDNA insert. Libraries
were constructed by Sugano et al. (University of Tokyo, Institute of
Medical Science). Custom primer used for sequencing
( 5' end primer [CTTCTGCTCTAAAGTGGG];
3' end primer [CCAGTCGACGCTCGACACA] ).
Location/Qualifiers
1..2905
/organism="Macaca fascicularis"
/db_xref="taxon:9541"
/clone="QnpA-10050"
/sex="male"
/tissue_type="brain parietal lobe"
/clone_lib="macaque brain cDNA library QnpA"
/dev_stage="adult"
1323..1670
/codon_start=1
/product="hypothetical protein"
/protein_id="BAB12122.1"
/db_xref="GI:9929931"
/translation="MGTRPFAAQTLPVVASLVSPSRQRLQRQCLVPGTGLPA
SLVAVVSPSPPTAPGRURITMAPGCPGRLLQPLPTPSFQGLTWSVLPSCVLML
FALCKNRLMLVIL"
BASE COUNT 687 a 721 c 687 g 808 t 2 others

```

ORIGIN

```

Alignment Scores:
Pred. No.: 3,32e-27 Length: 2905
Score: 567.00 Matches: 108
Percent Similarity: 99.08% Conservative: 0
Best Local Similarity: 99.08% Mismatches: 1
Query Match: 10.09% Indels: 0
DB: 9 Gaps: 0

US-09-830-837-6 (1-1052) x AB047598 (1-2905)

QY 944 SerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLysValValLeuPro 963
|||||
Db 1886 AGTAACCTTTGGAACATCAGAAGCTACTCTCCATGACCTGACAAAGTGGTTATCCC 1945
|||||
QY 964 AsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSerGlyAlaTrp 983
|||||
Db 1946 AACTTTGATCAATCGCCCTCAAGTGAGGCCCTTGTCCCTCGAGAGAGCGGTGCTGG 2005
|||||
QY 984 AspIleProGlyGlyIleMetProGlyArgTyrAsnGlnValGlyGlnThrIlePro 1003
|||||
Db 2006 GACATTCCTGGAGGATCATGCTGCGCCCTACACAGAGGTGGCCAGACCATTCCT 2065
|||||
QY 1004 ValPheAlaPheLeuGlyAlaMetValValLeuAlaPhePheValValGlnIleAsnLys 1023
|||||
Db 2066 GTCTTCGCTCTCTGGAGCCATGGTGGTCTTCTTGTGTGACAAATCAACAAG 2125
|||||
QY 1024 AlaLysSerArgProLysArgLysProArgValLysArgProGlnLeuMetGlnGln 1043
|||||
Db 2126 GCCAAGAGAGGCCGAAGCGAGGAGGAGCCAGGCGGTGAAGCGCCGAGCTCATGCAGCAG 2185
|||||
QY 1044 ValHisProLysThrProSerVal 1052
|||||
Db 2186 GTGCACCCGCCAAAGCCCTTCCTGGTG 2212

RESULT 25
AC112476 195287 bp DNA linear HTG 17-JUL-2002
LOCUS Rattus norvegicus clone CH230-122D23, *** SEQUENCING IN PROGRESS
DEFINITION ***; 65 unordered pieces.
ACCESSION AC112476
VERSION AC112476.3 GI:21745561
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 195287)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,
Alsbrooks, S.L., Anaratunge, H.C., Are, J.R., Ayele, M., Banks, T.,
Barbata, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D.,
Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
Fallis, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
Gabis, A., Gao, J., Garcia, A., Garner, T., Garza, P., Gill, R.,
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J.,
Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
Jacobson, B., Jia, X., Johnson, R., Jolivet, S., Joudah, S.,
Karlsone, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
Kratovic, J., Kurishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,

```

Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Oguh, M., Okwuonu, G.,
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoohart, N., Sisson, I.,
 Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Tatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmami, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 Direct Submission
 Unpublished
 2 (bases 1 to 195287)
 Worley, K.C.
 Direct Submission
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 195287)
 Worley, K.C.
 Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20303366.

 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

 Project Information
 Center project name: GSSS
 Center clone name: CH230-12D23

 Summary Statistics
 Sequencing vector: Plasmid
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 137024 bases at least Q40
 Consensus quality: 143926 bases at least Q30
 Consensus quality: 150275 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 65 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1 1272: contig of 1272 bp in length
 * 1273: gap of unknown length
 * 1373 1372: contig of 1368 bp in length
 * 2741 2840: gap of unknown length
 * 2841 4272: contig of 1432 bp in length
 * 4273 4372: gap of unknown length
 * 4373 5666: contig of 1294 bp in length
 * 5667 7266: gap of unknown length
 * 7267 7366: contig of 1500 bp in length
 * 7367 8912: gap of unknown length
 * 8913 9012: contig of 1546 bp in length
 * 9013 10022: gap of unknown length
 * 10023 10122: contig of 1010 bp in length
 * 10123 11538: contig of 1416 bp in length
 * 11539 11638: gap of unknown length
 * 11639 12973: contig of 1335 bp in length
 * 12974 13073: gap of unknown length
 * 13074 14297: contig of 1224 bp in length
 * 14298 14397: gap of unknown length
 * 14398 15608: contig of 1211 bp in length
 * 15609 15708: gap of unknown length
 * 15709 17207: contig of 1499 bp in length
 * 17208 17307: gap of unknown length
 * 17308 18882: contig of 1575 bp in length
 * 18883 18982: gap of unknown length
 * 18983 20753: contig of 1771 bp in length
 * 20754 20853: gap of unknown length
 * 20854 21905: contig of 1052 bp in length
 * 21906 22005: gap of unknown length
 * 22006 23612: contig of 1607 bp in length
 * 23613 25358: gap of unknown length
 * 25359 25459: gap of unknown length
 * 25460 27977: contig of 2519 bp in length
 * 27978 28077: gap of unknown length
 * 28078 29889: contig of 1812 bp in length
 * 29890 29989: gap of unknown length
 * 29990 31581: contig of 1591 bp in length
 * 31582 33566: gap of unknown length
 * 33567 33667: gap of unknown length
 * 33668 35307: contig of 1641 bp in length
 * 35308 35407: gap of unknown length
 * 35408 37763: contig of 2356 bp in length
 * 37764 37863: gap of unknown length
 * 37864 38877: contig of 1014 bp in length
 * 38878 38977: gap of unknown length
 * 38978 41067: contig of 2090 bp in length
 * 41068 41167: gap of unknown length
 * 41168 43639: contig of 2472 bp in length
 * 43640 43739: gap of unknown length
 * 43740 46078: contig of 2339 bp in length
 * 46079 46178: gap of unknown length
 * 46179 48378: contig of 2200 bp in length
 * 48379 48478: gap of unknown length
 * 48479 50170: contig of 1692 bp in length
 * 50171 50270: gap of unknown length
 * 50271 53202: contig of 2332 bp in length
 * 53203 53302: gap of unknown length
 * 53303 56020: contig of 2718 bp in length
 * 56021 56120: gap of unknown length
 * 56121 59634: contig of 3334 bp in length
 * 59635 59734: gap of unknown length
 * 59735 61539: contig of 1805 bp in length
 * 61540 61639: gap of unknown length
 * 61640 63419: contig of 1760 bp in length
 * 63420 63519: gap of unknown length
 * 63520 66012: contig of 2493 bp in length
 * 66013 66112: gap of unknown length
 * 66113 69383: contig of 3271 bp in length
 * 69384 69483: gap of unknown length
 * 69484 72141: contig of 2658 bp in length
 * 72142 72241: gap of unknown length
 * 72242 74600: contig of 2359 bp in length
 * 74601 74700: gap of unknown length
 * 74701 78269: contig of 3569 bp in length
 * 78270 78369: gap of unknown length
 * 78370 82451: contig of 4082 bp in length
 * 82452 82551: gap of unknown length
 * 82552 86231: contig of 3740 bp in length
 * 86232 86331: gap of unknown length
 * 86332 88846: contig of 2455 bp in length
 * 88847 88946: gap of unknown length
 * 88947 92247: contig of 3300 bp in length
 * 92248 92347: gap of unknown length
 * 92348 95087: contig of 2741 bp in length
 * 95088 95187: gap of unknown length

THIS PAGE BLANK (03/70)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 2, 2003, 01:36:29 ; Search time 3332 Seconds
(without alignments)
5113.341 Million cell updates/sec

Title: US-09-830-837-6

Perfect score: 5617

Sequence: 1 MRLVNIWLLVLLCGKKH.....PRVKRPMQOVHPKTPSV 1052

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlh
-Q/cgn2_1/USFTO.spool/US09830837/runat_23052003.181924.8071/app_query.fasta_1.1223
-DB=EST -QPM=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=25 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09830837 @CGN 1.1 2475 @runat_23052003.181924.8071 -NCPG=6 -ICPU=3
-NO_XLPAY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPOP=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vit:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1563	27.8	927	14	BQ921443
2	1510	26.9	887	14	BQ879057
3	1470	26.2	885	9	AU140223
4	1440	25.6	1047	13	BM552158
5	1434	25.5	918	14	BQ929998
6	1418	25.2	949	13	BI905658
7	1410.5	25.1	928	14	BQ955268
8	1399.5	24.9	838	14	BQ228507
9	1380	24.6	795	13	BI870081
10	1380	24.6	870	14	BQ217956
11	1376.5	24.5	971	13	BM475146
12	1372	24.4	792	14	BQ573358
13	1349.5	24.0	944	13	BI904898
14	1349	24.0	859	14	BQ688754
15	1348	24.0	858	13	BI084430
16	1340	23.9	766	13	BI914542
17	1319	23.5	812	10	BE386747
18	1295.5	23.1	929	13	BI414585
19	1294	23.0	975	14	BQ957796
20	1293.5	23.0	859	14	BQ928302
21	1291	23.0	736	12	BG697188
22	1289	22.9	733	14	BM949214
23	1279.5	22.8	809	9	AU139053
24	1254	22.3	724	13	BI661522
25	1248	22.2	880	13	BI091766
26	1244	22.1	741	14	BM945034
27	1237	22.0	965	12	BG482068
28	1232	21.9	783	13	BI455980
29	1231	21.9	721	9	AU133912
30	1222	21.8	862	13	BI413549
31	1217	21.7	698	12	BG695170
32	1210	21.5	729	14	BM963520
33	1209	21.5	666	14	BM783543
34	1209	21.5	666	14	BM792735
35	1200.5	21.4	1010	14	BQ689579
36	1192.5	21.2	804	9	AU119413
37	1181	21.0	647	9	AL046669
38	1180	21.0	652	14	BM829953
39	1174	20.9	691	9	AU118931
40	1165	20.7	688	9	AU123483
41	1164.5	20.7	960	12	BG023875
42	1163	20.7	697	10	BE313242
43	1156	20.6	954	14	BQ645116
44	1149	20.5	835	14	BQ443546
45	1147	20.4	840	13	BI905329

ALIGNMENTS

RESULT 1
BQ921443
LOCUS
DEFINITION AGNCOURT_8930283 NIH_MGC_94 Mus musculus cDNA clone IMAGE:6466228
5', mRNA sequence.
ACCESSION BQ921443
VERSION BQ921443.1
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 927)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLM13990 row: d column: 05
High quality sequence stop: 697.
Location/Qualifiers
1. 927
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:646628"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
227 a 236 c 251 g 212 t 1 others

BASE COUNT
ORIGIN

Alignment Scores:

Pred. No.: 3.77e-147 Length: 927
Score: 1563.00 Matches: 252
Percent Similarity: 97.07% Conservative: 6
Best Local Similarity: 95.11% Mismatches: 8
Query Match: 27.83% Indels: 1
DB: 14 Gaps: 0

US-09-830-837-6 (1-1053) x BQ921443 (1-927)

QY 677 ValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetVal 696
Db 3 GTGCTGGCGGCCCATTCACATGTTTGACCGCCACAGATATGGCACTTCTGCTGGTG 62
QY 697 AspSerGluGluTyrPheProGluGluLeuAlaLysLeuArgAspValAspAsn 716
Db 63 GACAGTGAGGAGAGTACTTCCCTGAGGAGATGCTAAGCTGAGGAGGATGTGACAA 122
QY 717 GlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerValMetArgIleValLys 736
Db 123 GGCCTTTCCTCGTCATCTTCAGTACTGGTACACACTTCGTATAGAAAGTGAAG 182
QY 737 PheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyClyAlaAsnIlePro 756
Db 183 TTTTATGATGAACACCCAGGAGGTGGTGGATGCCACACCGGAGGAGGACATCCCA 242
QY 757 AlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGly 776
Db 243 GCTCTGAATGAGCTGCTGTGTGGACATGGGGTTCAGTGACGGCTATATGAAGG 302
QY 777 GluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPhe 796
Db 303 GAGTTGTCTCGCAACCATGACATGATGACTATGCTGGGTGCGGGTGCAGCATGCCAAGTTT 362
QY 797 ProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLys 816
Db 363 CCAGAAGATGGCGTGTGATCACACAGACTTTCAAGGACCAAGGATGGAGGTCTTAAAA 422
QY 817 GlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGlu 836
Db 423 CAAGAGACAGCAGTTGTGGAATATGTTCCCATTTGGGGCTTTACAGATTCATCTGAA 482
QY 837 GlyClyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHisArgGln 856
Db 483 GTGTGGAGCGCGATGCTGTATGGAGACTCCCACTGCTTGGATGACATGACAGCAGAC 542

QY 857 LysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrPro 876
Db 543 AAGGACTGCTTTGGTTCCTGGATCGCTCTTCAGTACATCATATGGGTGACCCCT 602
QY 877 ProSerLeuSerHisSerGlyAsnArgGlnArgProSerClyAlaClySerValThr 896
Db 603 CCCAGCCCTCAGCCATCAGGAACCGGACGCCACCTAGCCAGGACCGGCTTGGCCCT 662
QY 897 ProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeu 916
Db 663 CCTGAAGATGGAAGAAACCACTCCATCGTACTTCCAAAGTCTTGAAGCCCACTTG 722
QY 917 GlyAspProLysProArgProLeuProAlaCysProArgLeuSerTrpAlaLysProGln 936
Db 723 GGAGACCCGAAACCTCGGCCCTGCGCAGCTGTCCACATTTGTTCATGGCCCAAGCCACAG 782
QY 937 -ProLeuAsnGluThrAlaProSerAsnLeuTrpLysHisGlnLysLeuSerIleAs 956
Db 783 CCTTTGAATGAGACGACCCCAAGTAATCTTTTGGAAATCAGAACATCAGAGCTCTCCATTGA 842
QY 956 pLeuAspLysValValLeuProAspPheArgSerAsnArgProGlnValArgProLeuSe 976
Db 843 CCTGGACAAAGTAGTGTGTACCAACTTTTCGATCCCAATCGGCTCAAGTGAGAACCTTGT 902
QY 976 rProGlyGluSerGlyAla 982
Db 903 CCTTGAAGAGAGTGTGCTGCC 921

RESULT 2

BQ879057

LOCUS

DEFINITION

AGENCOURT_8074363 NIH_MGC_110 Homo sapiens cDNA clone IMAGE:6085736

5', mRNA sequence.

ACCESSION

BQ879057

KEYWORDS

EST.

SOURCE

ORGANISM

human.

REFERENCE

1 (bases 1 to 887)

AUTHORS

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LMC2319 row: j column: 09

High quality sequence stop: 697.

Location/Qualifiers

1. 887

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6085736"

/clone_lib="NIH_MGC_110"

/tissue_type="ductal carcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pORF7; Site_1: XhoI;

Site_2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(C). Library constructed by

Ling Hong in the laboratory of Gerald M. Rubin (University

of California, Berkeley) using ZAP-CDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH_MGC Library."

BASE COUNT

203 a 231 c 237 g 215 t 1 others

QY 161 ProLeuArgAlaSerLeuSerLeuGlySerGlyPheTrpHisAlaThrGlyArgHis 180
 DB 361 CCCTCCGAAGAGCAGCTCTCCCTGGGCTCTGGCTTCTGGCATGCTACGGAAGGCAT 420
 QY 181 SerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAlaAsp 200
 DB 421 TCAGCAGACGCTGCTGAGAGCCATCCGCGCAGGTTGCCAGACACTGCAGGCAGAT 480
 QY 201 ValLeuTrpGlnMetGlyThrGlyAlaAsnValArgValAlaValPheAspThrGly 220
 DB 481 GTCTCTGCGAGATGGATATACAGTCTCAATGTAAGATGTCGTTTGTACATGGG 540
 QY 221 LeuSerGlyLysHisProHisPheLysAsnValLysGluArgThrAsnTrpHisGlu 240
 DB 541 CTGAGCGAGACATCCCACTTCAAAATGTGAGGAGAACCACTGACCAACGAG 600
 QY 241 ArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSerMet 260
 DB 601 CGAACCTGGAGATGGTGGGTCATGTCACATGTCGTGGCAGGTGTAGCCAGCATG 660
 QY 261 ArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThrAsn 280
 DB 661 AGGAGTCCCAAGGATGCTCCAGATCGACAACCTTCAATTTTCAGGGTCTTTACCAAT 720
 QY 281 AsnGlnValSerTrpTrpPheLeuAspAlaPheAsnTrpAlaIleLeuLysLys 300
 DB 721 AATCAGGATCTTACATCTTGGGTTTTTGGCGCGCTTAACTATGCCATTTTAAAGAA 780
 QY 300 sIleAspValLeuAsnLeuSerIleGlyProAspPheMetAspHis-PropheVala 320
 DB 781 GATCAGACGTGTAACCTCAGCATTCGCGGCCGAGCTTATGATCATCTCCGTGGTTG 840
 QY 320 sPlysValTrpGluLeuThrAlaAsnValIleMetValSer 334
 DB 841 ACCANGGCTGGGAANTACAGCTTACATGGATCATCGGTTCT 884

 RESULT 4
 BM552158
 LOCUS
 DEFINITION 1047 bp mRNA linear EST 20-FEB-2002
 5' mRNA sequence.
 ACCESSION BM552158
 VERSION BM552158.1 GI:18789785
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 1047)
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs@mail.nih.gov
 Tissue procurement: ATCC
 cDNA Library Preparation: Rubin Laboratory
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: LLM01994 row: a column: 14
 High quality sequence stop: 605.
 Location/Qualifiers
 1..1047
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:5476885"
 /clone_11b="NIH_MGC_98"
 /tissue_type="astrocytoma grade IV, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:

ECORI: cDNA made by oligo-dT priming. Directionally
 cloned into EcoRI/XhoI sites using the following 5'
 adaptor: GGCACGAG(G). Library constructed by Ling Hong
 in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

BASE COUNT 247 a 265 c 305 g 230 t
 ORIGIN

Alignment Scores:
 Pred. No.: 136e-134 Length: 1047
 Score: 1440.00 Matches: 291
 Percent Similarity: 87.17% Conservative: 8
 Best Local Similarity: 84.84% Mismatches: 29
 Query Match: 25.64% Indels: 15
 DB: 13 Gaps: 8

US-09-830-837-6 (1-1052) x BM552158 (1-1047)

QY 673 TyrPheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThr 692
 DB 1 TACTTTGTAGAGGTCTCGGGGCCCTTCACGTGTTTGTATGCCAGTCAGTATGGCACT 60
 QY 693 LeuLeuMetValAspSerGluGluGlyTyrPheProGluGluIleAlaLysLeuArgArg 712
 DB 61 TTGCTGATGGTGGACAGTGAAGGAGGAGTACTTCCCTGAAGAGATCGCAAGCTCCGGAGG 120
 QY 713 AspValAspAsnGlyLeuSerLeuValIlePheSerAspTrpTrpAsnThrSerValMet 732
 DB 121 GAGCTGGACAACGCCCTCTCGCTCGTCATCTTCAGTGACTGGTACAACTCTCTGTATG 180
 QY 733 ArgLysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGly 752
 DB 181 AGAAAGTGAAGTTTATGATGAAACACACAGGAGTGGTGGATGCCGATACCGGAGGA 240
 QY 753 AlaAsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGly 772
 DB 241 GCTACATCCAGCTCTGAATGAGCTGCTGTGTGGAACTATGGGTTCAGCATGGC 300
 QY 773 LeuTyrGluGlyLysPheThrLeuAlaAsnHisAspMetTyrTrpAlaSerGlyCysSer 792
 DB 301 CTGTATGAAGGGAGTTTACCCCTGGCCAACTACATGATGATATATCGTCAGGGTGACG 360
 QY 793 IleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGlyLeu 812
 DB 361 ATCCGGAAGTTTCCAGAAGATGGCTGCTGATACACAGACTTTTCAAGGACCAAGGATG 420
 QY 813 GluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGln 832
 DB 421 GAGGTTTAAAGCAGGAACACAGCAGTTGTTGAAAACGTCCTCCATTTTGGGACTTTATCAG 480
 QY 833 IleProAlaGluGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAsp 852
 DB 481 ATTCAGCTGAGGGTGGAGCGCGGATTTGCTGATGGGACTTCCAATTTGCTTGGATGAC 540
 QY 853 SerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTrpSerTyr 872
 DB 541 ACTCACACAGAGGACTGCTTTGGCTTCTGGATGCCCTCTCCAGTACACATCGAT 600
 QY 873 GlyValTrpProProSerLeuSerHisSerGlyAsnArgGlnArgProSerGlyVala 892
 DB 601 GGGGTGACACCGCTAGCTTCACTCTGGAAACCGCCAGCGCCCTCCAGTGGAGCA 660
 QY 893 GlySerValThrProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeu 912
 DB 661 GGCTCAGTCACTCCAGAGAGATGGAAGGAACCATCTTCTCGGTACTTCCAAAGGGTCTG 720
 QY 913 GluAlaHisLeuGlyAspProLysProArgProLeuProAlaCysPro-ArgLeuSerTr 932
 DB 721 GGAGGCCATTTGGGAGGAGGCCAAACCTCGGCTCTTACACGCTGTGCCAGCTGTCTGT 780
 QY 932 pAlaLysProGlnProLeu---AsnGluThrAlaProSerAsnLeuTrp-LysHisGlnL 951


```

Db      781  GGGCCAGCCACAGCCCTTTAAACGAGAGCGGGGCCAGGTAACCTTTGGGAAACATCAGA 840
Qy      951  ys---LeuLeuSerIleAsp-LysVal---ValLeuProAsnPheArgSer--- 967
Db      841  AGCTACTCTCCATTGACCTCGACAAAGGTGGGGTTAAACCCCAATTTTTCGAATCCG 900
Qy      968  AsnArgProGln-----ValArgProLeuSerPro---GlyGluSerGlyAlaTrpAsp 984
Db      901  AATCGCCCTCCCAAGGTGAAGGCCCTTGGTCCCTCGGAAAGAGCGCGGC--- 957
Qy      985  IleProGlyGlyIleMetProGlyArgTyrAsnGluValGlyGlnThrIleProVal 1004
Db      958  ---CCTGGGAACATTTCCTGGGAAGGAACATGTCGCTGGG-----CCGCC 1005
Qy      1005  Phe 1005
Db      1006  TTT 1008

RESULT 5
BQ929998
LOCUS   BQ929998
DEFINITION BQ929998 918 bp mRNA linear EST 20-AUG-2002
5', mRNA sequence.
AGENCY BQ929998
VERSION BQ929998.1 GI:22345029
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 918)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@email.nih.gov
Tissue procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM13998 row: g column: 18
High quality sequence stop: 646.

FEATURES
source
1..918
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:6469385"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 241 a 210 c 241 g 223 t 3 others
ORIGIN

Alignment Scores:
Pred. No.: 4.34e-134 Length: 918
Score: 1434.00 Matches: 287
Percent Similarity: 95.15% Conservative: 7
Best Local Similarity: 92.88% Mismatches: 6
Query Match: 25.53% Indels: 9
DB: 14 Gaps: 0

US-09-830-837-6 (1-1052) x BQ929998 (1-918)
Qy      40  HisLeuThrLeuLysValGluPheSerSerThrValValGluTyrGluTyrIleValAla 59

```

```

Db      2  CACCTGACCTTTGAGGTGGAAATTCCTCTCAACTGTGTGGAGTACGAATATATTGTGGCT 61
Qy      60  PheAsnGlyTyrPheThrAlaLysAlaArgAsnSerPheIleSerSerAlaLeuLysSer 79
Db      62  TTCACGGATACCTTCAGCCCAAGCTAGAAATCATTTATTCAAGTGTCTGAAAGC 121
Qy      80  SerGluValAspAsnTrpArgIleIleProArgAsnAsnProSerSerAspTyrProSer 99
Db      122  AGTGAAGTGGAAACATGAGAAATAATACCTCGGAACAACCCATCCAGTACTACCCTAGT 181
Qy      100  AspPheGluValIleGlnIleLysGluLysGlnLysAlaGlyLeuLeuThrLeuGluAsp 119
Db      182  GATTTTGGGTGATTCAGATAAAGAGAGACAGAGCGGGGTGCTTCACACTTGAAGAT 241
Qy      120  HisProAsnIleLysArgValThrProGlnArgLysValPheArgSerLeuLysTyrAla 139
Db      242  CACCCCAACATCAAGCGGTGACACCCACCGGNAAGTCTTTCGTCCTCAAGTTTGTCT 301
Qy      140  GluSerAspProThrValProCysAsnGluThrArgTrpSerGlnLysTrpGlnSerSer 159
Db      302  GAATCAACACCCCATCGTCCCTGTATGAAACCCGCTGGAGCCAGAGAGTGGCAGTCATCA 361
Qy      160  ArgProLeuArgArgAlaSerLeuSerLeuGlySerGlyPheThrPheHisAlaThrGlyArg 179
Db      362  CGTCCCTGAAAGAGCCAGTCTCTCCCTGGGCTCTGGATTCTGGCATGCAACAGGAAGA 421
Qy      180  HisSerSerArgArgLeuLeuArgAlaIleProArgGlnValAlaGlnThrLeuGlnAla 199
Db      422  CATTCAAGTCGGGATGTCTGAGAGCCATTCCTCGCCAGGTGCGCCAGACACTGCAGGCA 481
Qy      200  AspValLeuTrpGlnMetGlyTyrThrGlyAlaAsnValArgValAlaValAlaPheAspThr 219
Db      482  GATGTGCTGTGGCAGATGGGATACACAGGTGCTAATGTCAGAGTGTGCTTTTGTGATACT 541
Qy      220  GlyLeuSerGluLysHisProHisPheLysAsnValLysGluArgThrAsnTrpThrAsn 239
Db      542  GGGCTCAGTGAGAGCATCCGCATTTTGAAGATGTGAAGGAGAGAACCACTGGACCAAT 601
Qy      240  GluArgThrLeuAspAspGlyLeuGlyHisGlyThrPheValAlaGlyValIleAlaSer 259
Db      602  GAGCGGACCTCGATGATGGGTAGGCCATGGCACATTCGTCGAGGTGTGATGCGCAGC 661
Qy      260  MetArgGluCysGlnGlyPheAlaProAspAlaGluLeuHisIlePheArgValPheThr 279
Db      662  ATGAGGAGTGCCCAAGGATTTGCTCCAGATGCAGAGCTGCACATCTTCAGGGTCTTTACA 721
Qy      280  AsnAsnGlnValSerTyrThrSerTrp-PheLeuAspAlaPheAsnTyr-AlaIleLeuL 299
Db      722  AACAATCAGGTGTCTTACACATCTTGGGTTTCTGGATGCGCTTCAACTATGCCATCCTAA 781
Qy      299  YsLysIleAspValLeuAsnLeu-SerIleGlyGlyProAspPheMet-AspHisPropH 318
Db      782  AGAAGATGAGCGTCTCAACCTTAGCATCGTGGGGCCCGCACTCATGGATCATCCGTT 841
Qy      318  e-ValAspLysValTrpGluLeuThrAla-AsnAsnValIleMetValSerAlaIleGly 337
Db      842  TGGTGGACAAGGTGTGGGATTAACAGCTNACATTTGCTAATTATGGGTTCTGCTATTGGG 901
Qy      338  AsnAspGlyProLeu 342
Db      902  CATGATGGGACTCTC 916

RESULT 6
LOCUS   BQ905658
DEFINITION BQ905658 949 bp mRNA linear EST 16-OCT-2001
603167672F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5256040 5',
mRNA sequence.
ACCESSION BQ905658
VERSION BQ905658.1 GI:16168219
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

```


FEATURES

source

Location/Qualifiers
 1. .928
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:6204905"
 /clone_lib="Lupski_sciatic_nerve"
 /sex="male"
 /tissue_type="sciatic nerve"
 /dev_stage="adult, 70 yr"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1: NotI; Site_2: SalI; cDNA made by oligo-dr priming. Directionally cloned using the following adaptors: 5'-TCGACCCAGCGTCG-3' and 5'-GACTAGTTCAGTCGCGAGCGCCCT(15)-3'. Size selected > 1 kb for average insert length 1.87 kb. This is a primary library, non-amplified. Library constructed by Life technologies and donated by J. Lupski, M.D./Ph.D. (Baylor College of Medicine) and is available through Life Technologies."

BASE COUNT 226 a 228 c 258 g 211 t 5 others
 ORIGIN

Alignment Scores:

Pred. No.: 1.06e-131 Length: 928
 Score: 1410.50 Matches: 287
 Percent Similarity: 92.90% Conservative: 1
 Best Local Similarity: 92.58% Mismatches: 9
 Query Match: 25.11% Indels: 13
 DB: 14 Gaps: 3

US-09-830-837-6 (1-1052) x BQ955268 (1-928)

QY 647 AsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPhaArgAspMetTyrGln 666
 Db 2 AATGACCCCTTACACTGGAGTGGTATCATCATCCACCAATTCAGGAGATATGACCAAG 61
 QY 667 HisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThrCysPheAsp 686
 Db 62 CATCTGAGAAGCATGGCTACTTGTAGAGGTCTCGGGGCCCTTCACGTTTGTAT 121
 QY 687 AlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluTyrPheProGluGlu 706
 Db 122 GCACATCATATGGCACTTGTGATGGTGGACAGTGGAGAGTACTTCCTCGAAGAG 181
 QY 707 IleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValIlePheSerAspTyr 726
 Db 182 ATCGCAACTCCGGAGGACGTGGACAACGGCCCTCTCGCTCATCTTCAGTACTGG 241
 QY 727 TyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpTyr 746
 Db 242 TACAACACTTCTGTATGAGAAAGTGAAGTTTATGATGAAACACAGGAGTGTGG 301
 QY 747 MetProAspThrGlyClyAlaAsnIleProAlaLeuAsnGluLeuSerValTrpAsn 766
 Db 302 ATGCCGGATACCGGAGGACGTAAATCCAGCTCCAGCTGCTGTCTGTGTGAAC 361
 QY 767 MetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyr 786
 Db 362 ATGGGGTTCAGGATGGCTGTATGAAGGGAGTTCACCTGGCCCAACCATGACATGTAT 421
 QY 787 TyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIleThrGlnThr 806
 Db 422 TATGGTCCAGGTGAGCATCCGGAAGTTTCAGAGATGGCGTGGTATACACAGACT 481
 QY 807 PheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValValGluAsnValPro 826
 Db 482 TTCAAGGACCAAGGATGGAGTTTAAAGCAGGAAACAGCAGTGTGTGAAACGTCGCC 541
 QY 827 IleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleValLeuTyrGlyAsp 846
 Db 542 ATTTGGGACTTTATCAGATCCAGTGAAGGTGGAGCGCGATTGTACTGTATGGGAC 601

QY 847 SerAsnCysLeuAspAspSerHisArgGlnLysAspCysPheTrpLeuLeuAspAlaLeu 866
 Db 602 TCCAATTGCTTGGATGACAGTACCACAGAGGACTGCTTTTGGCTTCTGGATGCCCTC 661
 QY 867 LeuGlnTyr-ThrSerTyrGlyValThrProProSerLeuSerHisSerGlyAsnArgG 886
 Db 662 CTCCAGTACATCATGATGGGTGACACCGCTAGCCTCAGTCACTCTGGGAACGCCA 721
 QY 886 nArgProProSerGlyAlaGlySerValThrProGlu-ArgMetGlu-GlyAsnHisLeu 905
 Db 722 GGGCCCTCCAGTGGAGGAGTCACTCACTCCAGAGAGAGTGAAGAAAGAAACCATCTT 791
 QY 906 HisArgTyr-SerLys-ValLeuGluAlaHis----LeuGlyAspPro-LysProArg-Pr 923
 Db 782 CATCGGTACTCCCAAGGTTCTGGNANGCCCATTTTGGGAGACCCCAAAACCTCGNCC 841
 QY 923 OleuPro-AlaCys-ProArgLeuSer---Trp-AlaLysProGlnProLeu---AsnG 940
 Db 842 TCTACCAACNCTGTTCCACGGCTTGTCTGGGGCCAGGCCACAGCCCTTTAAACAGAG 901
 QY 940 uThrAlaPro 943
 Db 902 AACGGCGCCC 911

RESULT 8

BQ228507

LOCUS

DEFINITION

5', mRNA sequence.

ACCESSION

BQ228507

VERSION

BQ228507.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 838)

AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL

Unpublished (1999)

COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs@email.nih.gov

Tissue Procurement: DCTD/DTP/Gazdar

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM13355 row: d column: 14

High quality sequence stop: 606.

FEATURES

Location/Qualifiers

1. .838

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6070621"

/clone_lib="NIH_MGC_68"

/tissue_type="large cell carcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: oligo dr.

Average insert size 1.8 kb. Library constructed by Life

Technologies."

BASE COUNT 214 a 217 c 213 g 193 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.15e-130 Length: 838
 Score: 1399.50 Matches: 272
 Percent Similarity: 83.28% Conservative: 2
 Best Local Similarity: 82.67% Mismatches: 4
 Query Match: 24.92% Indels: 52
 DB: 14 Gaps: 2

US-09-830-837-6 (1-1052) x BQ228507 (1-838)

LOCUS BI870081 795 bp mRNA linear EST 11-OCT-2000
DEFINITION 60339350F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:5403972 5',
mRNA sequence.
ACCESSION BI870081
VERSION BI870081.1 GI:16043754
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE NIH-MGC http://mgc.ncl.nih.gov/
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: ATCC
CDNA library Preparation: Life Technologies, Inc.
DNA Sequencing by: InCyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLNL2030 Row: c. column: 13
High quality sequence stop: 794.
Location/Qualifiers
1. 795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5403972"
/clone_lib="NIH_MGC_90"
/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 203 a 196 c 206 g 190 t
ORIGIN
Alignment Scores:
Pred. No.: 9,86e-129 Length: 795
Score: 1380.00 Matches: 260
Percent Similarity: 98.86% Conservative: 1
Best Local Similarity: 98.48% Mismatches: 3
Query Match: 24.57% Indels: 0
DB: 13 Gaps: 0
US-09-830-837-6 (1-1052) x BI870081 (1-795)
QY 113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAGCCAGCCCTCTCCCTGGGCTCTGGC 181
QY 173 PheTrpHisAlaThrGlyArgHisSerSerArgArgLeuLeuLeuArgAlaIleProArgGln 192
182 TTCTGGCATGCTACGGGAAGGCATTCGAGCAGACGGCTGCTGAGAGCCATCCCGCCGCG 241
QY 193 ValAlaGlnThrLeuGlnAlaAspValLeuTrpGlnMetGlyTrpThrGlyAlaAsnVal 212
242 GTTCCCAACACTGACGGCAGATGTCTCTGGCAGATGGGATATACAGTCTCTAATGTA 301
QY 213 ArgValAlaValPheAspThrGlyLeuSerGluLysHisProHisPheLysAsnValLys 232
113 GlyLeuThrLeuGluAspHisProAsnIleLysArgValThrProGlnArgLysVal 132
2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
Db 2 GGCGTCTAAACACTGAAGATCATCCAAACATCAACGGGTCAACGCCCAACGAAAGATC 61
QY 133 PheArgSerLeuLysThrAlaGluSerAspProThrValProCysAsnGluThrArgTrp 152
62 TTTGCTTCCTCAAGTATGCTGATCTGACCCCAAGTACCTGCAATGAAACCCGGTGG 121
QY 153 SerGlnLysTrpGlnSerSerArgProLeuArgGArgAlaSerLeuSerLeuGlySerGly 172
122 AGCCAAAGTGGCAATCATCATGTCCTCCGAGAG

Db 302 AGAGTTCCTGCTTTTGGACACTGGCTGAGCGAGAGCATCCCACTTCAAAATGTGAAG 361

QY 233 GluArgThrAsnTrpThrAsnGluArgThrLeuAspAspGlyLeuGlyHisGlyThrPhe 252

Db 362 GAGAGAACCACTGGACCAACGAGCGACGCTGGAGATGGGTGGGCCATGGCACATTC 421

QY 253 ValAlaGlyValIleAlaSerMetArgGluCysGlnGlyPheAlaProAspAlaGluLeu 272

Db 422 GTGGCAGGTGTGATAGCCAGCATGAGGAGTGCACAGGATTTGCTCCAGATGCAAGATT 481

QY 273 HisIlePheArgValPheThrAsnAsnGlnValSerTyrThrSerTrpPheLeuAspAla 292

Db 482 CACATTTTCAGGGTCTTTTACCATAATCAGGTATCTTACACATCTTTGGTTTGGACGCC 541

QY 293 PheAsnTyrAlaIleLeuLysIleAspValLeuAsnLeuSerIleGlyGlyProAsp 312

Db 542 TTCACATATGCCATTTTAAAGAGATCGACGTGTAAACCTCAGCATCGCGGCCGGAC 601

QY 313 PheMetAspHisProPheValAspLysValTrpGluLeuThrAlaAsnValIleMet 332

Db 602 TTCATGGATCATCCGTTTGTGACAAAGTGTGGGAATTAAACAGCTAACCAATGTAATCATG 661

QY 333 ValSerAlaIleGlyAsnAspGlyProLeuTyrGlyThrLeuAsnAsnProAlaAspGln 352

Db 662 GTTTCGTCTATGGCAATGACGACCTCTTTATGGCACTCTGTAATTAACCTGCTGATCAA 721

QY 353 MetAspValIleGlyValGlyIleAspPheGluAspAsnIleAlaArgPheSerSer 372

Db 722 TGGATTGTGATGGAGTACGGCATTTGACTTTGAAGATAACATCGCCGCTTTCTTCA 781

QY 373 ArgGlyMetThr 376

Db 782 AGGGAGTGAAT 793

RESULT 10

LOCUS BQ217956 870 bp mRNA linear EST 02-MAY-2002

DEFINITION AGENCOURT_7546723 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6025653

ACCESSION BQ217956

VERSION BQ217956.1 GI:20399356

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 870)

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

plate: LLNL3238 row: b column: 22

High quality sequence stop: 709.

Location/Qualifiers

1..870

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:6025653"

/clone_lib="NIH_MGC_70"

/tissue_type="epithelioid carcinoma"

/lab_host="DH10B (phage-resistant)"

/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.1 kb. Library constructed by Life Technologies."

FEATURES

source

BASE COUNT 228 a 196 c 237 g 207 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 1.16e-128 Length: 870

Score: 1380.00 Matches: 265

Percent Similarity: 95.39% Conservative: 4

Best Local Similarity: 93.97% Mismatches: 8

Query Match: 24.57% Indels: 5

DB: 14 Gaps: 1

US-09-830-837-6 (1-1052) x BQ217956 (1-870)

QY 564 IleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHisVal 583

Db 2 ATCTCCATTCCTGTGACCAAGAAGCGCTCTCTGGGAAGGCATTGCTCAGGGCCATGTC 61

QY 584 MetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGlnThrSer 603

Db 62 ATGATCACTGTGGCTTCCCAGCAGACAGACAGTCAAAAATGGTGCAGAACACAGCTCA 121

QY 604 ThrValLysLeuProIleLysValLysIleIleProThrProProArgSerLysArgVal 623

Db 122 ACAGTAAAGCTCCCAATTAAGTGAAGATAATTCCTACTCCCCCGGAGCAAGAGAGTT 181

QY 624 LeuTrpAspGlnTyrHisAsnLeuArgTyrProProGlyTyrPheProArgAspAsnLeu 643

Db 182 CTCTGGGATCAGTACCACCAACCTCCGCTATCCACCTGGCTATTTCCACAGGATAATTTA 241

QY 644 ArgMetLysAsnAspProLeuAspTrpAsnGlyAspHisIleHisThrAsnPheArgAsp 663

Db 242 AGGATGAAGAATGACCCCTTTAGACTGGAATGGTGGATCAGATCCACACCACTTCAGGGAT 301

QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGlnValLeuGlyAlaProPheThr 683

Db 302 ATGTACAGCATCTGAGAAGCATGGGCTACTTTGTAGAGGTCTCTCGGGGCCCCCTTCACG 361

QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGlyThrPhe 703

Db 362 TGTTTGTATGCCAGTCAGTATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 421

QY 704 ProGluGluIleAlaLysLeuArgArgAspValAspAsnGlyLeuSerLeuValIlePhe 723

Db 422 CCTGAAGAGATCCCAAGCTCCGGAGGGAGCGTGACAAACGGCTCTCGCTCGTCATGTC 481

QY 724 SerAspTrpTyrAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArg 743

Db 482 AGTCACTGGTACCAACACTTCTGTTATGAGAAAAGTGAAGTTTATGATGAAAACACAGG 541

QY 744 GlnTrpTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLeuSer 763

Db 542 CAGTGTGGATGCGGATACCGGAGGAGCTAACATCCAGCTCTGAATGAGCTGCTGCTCT 601

QY 764 ValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHis 783

Db 602 GTGTGAACATGGGGTTCAGCATGGCTCTATGAAGGGGAGTTCACCTCGGCCAACCAT 661

QY 784 AspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIle 803

Db 662 GACATGTATATGCGTCCAGGTCAGCATCGCGAGTTTCCAGAAGATGGCGTCGTGATA 721

QY 804 ThrGlnThrPheLysAspGln-GlyLeuGluValLeuLys-----GlnGluThrAlaVal 821

Db 722 ACACAGACTTTTCAAGGACCAAGGATTTGGAGTTTANAGCAGGGAACACAGCAGCTGTGT 781

QY 821 lValGluAsnValProIleLeuGly--LeuTyrGlnIleProAlaGluGlyGlyArg 840

Db 782 TTGAAAAAAGCGTCCCATTTTGGGAACCTTTATCAGATTCCACGCTTGGAGGTGGGAAG 841

RESULT 11

BM475146 971 bp mRNA linear EST 05-FEB-2002

LOCUS BM475146

DEFINITION AGENCOURT_6477357 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5559245

```

5', mRNA sequence.
BM475146
VERSION BM475146.1 GI:18524188
EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 971)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Cloning Strategy: Agencourt Bioscience Corporation
Cloned Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LAM12283 row: 1 column: 06
High quality sequence stop: 650.
FEATURES
source
Location/Qualifiers
1..971
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5559245"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dt primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."
BASE COUNT 234 a 240 c 276 g 220 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 3,22e-128 Length: 971
Score: 1376.50 Matches: 287
Percent Similarity: 89.78% Conservative: 3
Best Local Similarity: 88.85% Mismatches: 17
Query Match: 24.51% Indels: 16
Dbs: 13 Gaps: 6
US-09-830-837-6 (1-1052) x BM475146 (1-971)
QY 664 MetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAlaProPheThr 683
DB 5 ATGTACCAGCATCTGAGAACGATGGCTACTTTGTAGAGGTCTCTCGGGGCCCTTCACG 64
QY 684 CysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluGluGluTyrPhe 703
DB 65 TGTGTTGATGCCAGTCAGTATGGCACTTCTGCTGATGGTGGACAGTGGAGGAGTACTTC 124
QY 704 ProGluGluIleAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValIlePhe 723
DB 125 CTGAAGAGATGCCAAGCTCGGAGGAGCTGTGACACAGCGCTCTCGCTGCTCATCTTC 184
QY 724 SerAspTyrPheAsnThrSerValMetArgLysValLysPheTyrAspGluAsnThrArg 743
DB 185 ACTGACTGTGACACACTTCGTGTATGAGAAAGTGAAGTATGATGAAACACACAGG 244
QY 744 GlnTyrPheMetProAspThrGlyAlaAsnIleProAlaLeuAsnGluLeuLeuSer 763
DB 245 CAGTGTGGATGCGGATACCGAGGAGCTACATCCAGCTCTGAATGAGTGTCTGTCT 304
QY 764 ValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAsnHis 783
DB 305 GTGTGAACATGGGGTTCAGCGATGGCTGTATGAGAGGGAGTTCACCCCTGGCCCAACAT 364

```

```

784 AspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValIle 803
DB 365 GACATGTATTATCGTCAGGTGCAGCATCGGAAGTTTCCAGAAGATGGCTGCTGATA 424
QY 804 ThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValAlaGlu 823
DB 425 ACACAGACTTTTCAAGGACCAAGATTGGAGCTTTTAAAGCAGGAAACAGCAGTGTGTGAA 484
QY 824 AsnValProIleLeuGlyLeuTyrGlnIleProAlaGlyCylGlyValGluValLeu 843
DB 485 AACGTCCCATTTTGGGACTTTATCAGATTCCAGTGGGTGGAGCGCGGATTTACTG 544
QY 844 TyrGlyAspSerAsnGlyCysLeuAspSerHisArgGlnLysAspCysPheThrLeuLeu 863
DB 545 TATGGGAGCTCCAAATGCTTGGATGACAGTCCAGCAGGAGGAGCTTTTGGCTCTG 604
QY 864 AspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSerGly 883
DB 605 GATGCCCTCTCCAGTACACATCGTATGGGTGACACCGCTAGCCTCAGCTCAGTCTGGG 664
QY 884 AsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGlu-GlyAs 903
DB 665 AACCCGCCAGGCGCTCCAGTGGAGCAGCTCAGTCTCAGAGAGGATGGAAGGAAA 724
QY 903 nHisLeuHisArgTyr-SerLys-ValLeuGluAla-HisLeu-GlyAspProLysPro-- 921
DB 725 ACATCTTCATCGTACTCCAGGCTTCTGTGAGGCCCTTGTGGGAGACCCCAAAACCTT 784
QY 922 -ArgProLeuProAla-CysProArg-LeuSerTyr-AlaLysProGlnProLeu--As 939
DB 785 CGGGCTCTACCAAGCGCTGCCAGCGCTTGTCTTGGGCGCAAGCCAGCCGCTTTTAAAC 844
QY 939 nGluThr---AlaProSerAsnLeuTyr---LysHisGlnLys-----LeuLe 953
DB 845 GGAGACGGCGGCCCAAGTAACTTTTGGAAACATCAGAAAGCTACTCTCCCAATGA 904
QY 953 uSerIleAspLeuAspLysValValLeuProAsnPhe---ArgSerAsnArgPro 970
DB 905 ACCNFGGACACAGGGGGGTGTTAACCAACCTTTTCGAATCGATCGATCGGCC 959

```

```

RESULT 12
BQ573358 792 bp mRNA linear EST 19-JUN-2002
LOCUS UI-M-FD0-byg-b-18-0-UI.r1.NIH_BMAP_FD0 Mus musculus cDNA clone
DEFINITION IMAGE:5717681.5', mRNA sequence.
ACCESSION BQ573358
KEYWORDS BQ573358.1 GI:21476675
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 792)
NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Cloned Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pyx-5.
Location/Qualifiers
1..792
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"

```

```

FEATURES
source
Location/Qualifiers
1..792
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"

```

```

/clone="IMAGE:5717681"
/clone_lib="NIH_BMAP_FDO"
/tissue_type="whole brain"
/dev stage="embryo 12.5 dpc"
/lab host="DHI0B (T1 phage resistant)"
/note="Organ: brain; Vector: pfx-Asc; Site:1: EcoR I; Site:2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pfx-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TCAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 193 a 208 c 215 g 174 t 2 others
ORIGIN

Alignment Scores:
Pred. No.: 6.33e-128 Length: 792
Score: 1372.00 Matches: 251
Percent Similarity: 98.11% Conservative: 8
Best Local Similarity: 95.08% Mismatches: 5
Query Match: 24.43% Indels: 0
DB: 14 Gaps: 0

US-09-830-837-6 (1-1052) x BQ573358 (1-792)

Qy 514 ValAsnValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAspLysProAsp 533
Db 1 GTTAATGTCACCATCCATCGATGGATGGCGTCACAGGAAGAAATGTGGATAAGCCCTGAG 60

Qy 534 TrpGlnProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSer 553
Db 61 TGGCGACCCCTATTACACAGATGGAGACAACTTGAAGTGGCTTCCTACTCTCTCA 120

Qy 554 ValLeuTrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaAla 573
Db 121 GTGTGTGGCCCGTGGTCCAGTTACCTTGGCATCTCCATTCTGTGACCAAGAGCCAGCT 180

Qy 574 SerTyrGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThr 593
Db 181 TCCTGGGAAGCATCGCTCAGGCCACATCATGATCATGCGTCCCGCCAGCAGACCA 240

Qy 594 GluSerLysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIle 613
Db 241 GAGTTACACAGTGGTGGGAGCACACTTCCACCGTGAAGCTGCCCATCAAGTGAAGATC 300

Qy 614 IleProThrProProArgSerLysArgValLeuTrpAspGlnTyrHisAsnLeuArgTyr 633
Db 301 ATTCCCAACCCCTCTCCGAGCAAGAGAGTCTCTCGGACCACTACCAACACCTCCGCTAC 360

Qy 634 ProGlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTyrAsn 653
Db 361 CCACCTGGCACTTCCCGAGGACAACTTCGGATGAAGATGACCCCTTAGACTGGAAAT 420

Qy 654 GlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyr 673
Db 421 GGGGACCACTCCACACCACTTCAGGACATGTACCATGTCGCGACATGGCGCTAC 480

Qy 674 PheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeu 693
Db 481 TTCGTGGAGGTGTCGGCGGCCCATTCACATGTTTTCAGCCACACACAGTAGTGCACATTG 540

Qy 694 LeuMetValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArgArgAsp 713
Db 541 CTGCTGTGTGACAGTAGGGAAGAGTACTTCCCTCGAGGAGATTGCTAAGCTGAGGAGGAT 600

714 ValAspAsnGlyLeuSerLeuValIlePheSerAspTyrPyrAsnThrSerValMetArg 733
Db 601 GTGGACAATGGCCCTTTCCCTCGTCATCTTCAGTCAGCTGGTACACACTTCTGTTATGAGA 660

734 LysValLysPheTyrAspGluAsnThrArgGlnTrpMetProAspThrGlyGlyAla 753
Db 661 AAGTGAAGTTTATGATGAACACACAGGAGTGGTGGATGCCAGACACCGGAGGAGCG 720

754 AsnIleProAlaLeuAsnGluLeuSerValTrpAsnMetGlyPheSerAspGlyLeu 773
Db 721 AACATCCAGCTCGAATGAGCTGCTGTGTGTGNGACATGAGGGGGTCAGTGACGGNCTA 780

774 TyrGluGlyGlu 777
Db 781 TATGAAGGGGAG 792

RESULT 13
BI904898
LOCUS
DEFINITION 603168937F1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5257041 5', mRNA sequence. 944 bp mRNA linear EST 16-OCT-2001
ACCESSION BI904898
VERSION BI904898.1 GI:16167312
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 944)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: ccapbs-femail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Plate: LLAM1649 row: i column: 10
High quality sequence start: 36
High quality sequence stop: 926.
FEATURES
source
1..944
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:5257041"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DHI0B (phage-resistant)"
/note="Organ: lung; Vector: pT73b-Pac (Pharmacia) with a modified polylinker; Site:1: NotI; Site:2: EcoRI; 1st strand cDNA was prepared from mRNA obtained from pooled lung tumors with a Not I - oligo(dT) primer [5'-TGTTACCAATCTGAGGGAGGGCGGCCCTCTGTTTTTTTTTTTTTTT 3'-]. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 228 a 229 c 259 g 227 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 1.66e-125 Length: 944
Score: 1349.50 Matches: 273
Percent Similarity: 91.83% Conservative: 8
Best Local Similarity: 89.22% Mismatches: 21
Query Match: 24.03% Indels: 9

```


Db 301 GAGCTGGACAAAGCCCTCTCGCTGCATCTTCAGTGACTGGTACAAACACTTCTGTATG 360

QY 733 ArgLysValLysPheTyrAspGluAsnThrArg-GlnTyrTrpMetProAspThrGlyGlu 752

Db 361 AGAAAGTGAAGTTTATGATGAACACACAGGCGAGTGGTGTATGCCGGATACCGGAGG 420

QY 752 yAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspG1 772

Db 421 AGCTAAATCCCAAGCTCTGAATGAGCTGTGTGTGTGAACATGGGTTTCAGCATGG 480

QY 772 yLeuTyr-GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCys 792

Db 481 CCTGTATGAGGGGAGTTCACCTCGCCACCAACCATGACATGATATGCGTCAGGGTGCA 540

QY 792 erIleAlaLysPheProGluAspGlyValIleThrGlnThrPheLysAspGln-Gly 811

Db 541 GCATCCGGAAGTTTCCAGAAGATGGCGTGTGTGATAACACAGACTTTCAGGACCAAGGA 600

QY 812 LeuGluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyr 831

Db 601 TTGGAGGTTTANAGAGGAACACAGAGTGTGTGAACAGCTCCCATTTTGGGACTTTAT 660

QY 832 GlnIleProAlaGluGly-GlyArgIleValLeuTyrGlyAspSerAsnCysLeuAs 851

Db 661 CAGATCCAGCTGAGGGTGAAGCGGATGTACTGTATGGGAGTCCCAATTTGCTTGA 720

QY 851 pAspSerHisArgGlnLysAspCysPheTrp-LeuLeuAspAla-LeuLeuGlnTyrThr 870

Db 721 TGACATCCCCGACACAGAGACTGCTTTTGGGCTTCTGGAAGCCCTCTCCAGTACACA 780

QY 871 Ser-TyrGlyValThrPro-----ProSerLeuSerHisSerGlyAsnArgGlnArgPr 888

Db 781 TCGGTATGGGTGGACACCGCCCTAACCCTCCGTCCT--TCTGGGAACCGCCCAAGGCC 837

QY 888 o----ProSerGlyAla 892

Db 838 CTTCCCAATGGCAGC 853

RESULT 15
BI084430
LOCUS
DEFINITION 858 bp mRNA linear EST 20-JUN-2001
602859561F1 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:5014358 5',
mRNA sequence.
ACCESSION BI084430
VERSION BI084430.1 GI:14502760
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 858)
NIH-MGC <http://mgi.ncl.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
CONTACT: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: CLONTECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>
Plate: L1CM1821 row: 1 column: 15
High quality sequence stop: 839.
Location/Qualifiers
FEATURES
source
1..858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5014358"
/clone_lib="NIH_MGC_102"
/tissue_type="epidermoid carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"

/note="Organ: salivary gland; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GCCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies). Note: this is a NIH_MGC Library."

BASE COUNT 200 a 215 c 246 g 197 t
ORIGIN

Alignment Scores:

Pred. No.: 1,98e-125 Length: 858
Score: 1348.00 Matches: 268
Percent Similarity: 93.75% Conservative: 2
Best Local Similarity: 93.06% Mismatches: 14
Query Match: 24.00% Indels: 6
DB: 13 Gaps: 0

US-09-830-837-6 (1-1052) x BI084430 (1-858)

QY 736 LysPheTyrAspGluAsnThrArgGlnTyrTrpMetProAspThrGlyGlyAlaAsnIle 755

Db 3 AAGTTTATGATGAACACACAGCAGTGTGTGGATCCGGATACCGAGGAGCTAAACATC 62

QY 756 ProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGlu 775

Db 63 CCAGCTCTGAATGAGCTGTCTGTGTGGAACATGGGTTCCAGCATGGCCGTATGAA 122

QY 776 GlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAlaLys 795

Db 123 GGGGAGTTCACTCCCTGGCCAAACCATGACATGATATATCGTTCAGGGTTCAGCATCGCGAAG 182

QY 796 PheProGluAspGlyValIleThrGlnThrPheLysAspGlnGlyLeuGluValLeu 815

Db 183 TTTCCAGAGATGGCGTCGTGATAACACAGACTTCAAGGACCAAGGATTTGGAGGTTTAA 242

QY 816 LysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIleProAla 835

Db 243 AAGCAGGAACACAGACTGTGTGAAACGCTCCCATTTTGGGACTTTATCATGATCCAGCT 302

QY 836 GluGlyGlyGlyArg-IleValLeuTyrGlyAspSerAsnCysLeuAspSerHisAr 855

Db 303 GAGGGTGGAGCGCGCATTTGCTACTGTATGGGGATCCCAATTTGCTGGATGACATCACCG 362

QY 855 gGlnLysAspCysPheThrLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValTh 875

Db 363 ACAGAGGACTGCTTTTGGCTTCTGGATGCCCTCCCTCCATACATCGTATGGGGTGAC 422

QY 875 rProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGlyAlaGlySerVa 895

Db 423 ACCGCTAGCTCAGTCACCTCTGGAAACCGCGAGCGCCCTCCCGATGGAGCAGCTCAGT 482

QY 895 lThrProGluArgMetGluGlyAsnHisLeuHisArgTyrSerLysValLeuGluAlaH1 915

Db 483 CACTCCAGAGAGGATGAAGAAACATCTTCATCGGTACTCCAAGGTTCTGGAGGCCCA 542

QY 915 sLeuGlyAspProLysProArgProLeuProAlaCysProArgLeuSerTrpAlaLysPr 935

Db 543 TTTGGGAGACCCCAAGAGCTCGGCTCTTACCAGCGCTGTCCACGCTGTGCTTGGGCCAAGCC 602

QY 935 oGlnProLeuAsnGluThrAlaProSerAsnLeu-TrpLysHisGlnLysLeuLeuSerI 955

Db 603 ACAGCCTTTAAACGAGACGCGCGCCAGTAACCTTTTGGAAACATCAGAACCTTATCCCA 662

QY 955 leAspLeuAspLysValValLeuProAsnPheArgSerAsnArgProGlnValArgProL 975

Db 663 TTGACCTGGACAAAGTGGTGTATACCAAAATTTCCGATCGAATCGCCCTCAAGTAGGGCCCT 722

QY 975 euSerPro-GlyGluSerGlyAlaTrpAspIleProGly-GlyIleMetProGlyArgTy 994

Db 723 TGTCCTTGGAAAGAGAGCGCGCTGGGACATTCCTGGAGGGGATCATGCTCGGGGCGCTA 782

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3614338"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pOTB7; Site_1: xhoi; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT 197 a 187 c 240 g 188 t
ORIGIN

Alignment Scores:
Pred. No.: 1.55e-122 Length: 812
Score: 1319.00 Matches: 250
Percent Similarity: 98.82% Conservative: 1
Best Local Similarity: 98.43% Mismatches: 1
Query Match: 23.48% Indels: 2
DB: 10 Gaps: 0

US-09-830-837-6 (1-1052) x BE386747 (1-812)

Qy 663 AspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGlu-ValLeuGlyAlaProPh 682
Db 39 GATATGATACAGATCTGAGAACGATGGCTACTTTGTAGAGGTCTCCGGGCCCCCTT 98
Qy 682 eThrCysPheAspAlaSerGlnTyrGlyThrLeuMetValAspSerGluGluGluTy 702
Db 99 CACGTGTTTGTATGCCAGTCAGTATGCCACTTTCGTATGTTGGACAGTGGAGGAGTA 158
Qy 702 rPheProGluGluLeuAlaLysLeuArgAspValAspAsnGlyLeuSerLeuValII 722
Db 159 CTTCCCTGAGAGATCCCAAGCTCCGAGGAGCGTGGACAAGGCCCTCTCGTCTCAT 218
Qy 722 ePheSerAspTyrTrpAsnThrSerValMetArgLysValPheTyrAspGluAsnTh 742
Db 219 CTTCACTGATGTTGACAACTCTCTGTATGAGAAAGTGAAGTTTATGATGAAACAC 278
Qy 742 rArgGlnTrpMetProAspThrGlyGlyAlaAsnIleProAlaLeuAsnGluLeuLe 762
Db 279 AAGCACTGTGGATGCCGATACCGAGGAGCTAATCCAGCTCTGAATGAGCTGCT 338
Qy 762 uSerValTrpAsnMetGlyPheSerAspGlyLeuTyrGluGlyGluPheThrLeuAlaAs 782
Db 339 GTCTGTGTGAACATGGGGTTCACCGATGGCTGTATGAAGGGAGTTCACCCCTGGCAA 398
Qy 782 nHisAspMetTyrTrpAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVa 802
Db 399 CCATGACATGATATATCCGTCAGGTGCAGATCCGAAATTTCCAGAAGATGGCGCT 458
Qy 802 lIleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValVa 822
Db 459 GATAACACACACTTTCAGGACCAAGGATGGAGTTTAAACAGGGAACACAGATTGT 518
Qy 822 lGluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleVa 842
Db 519 TGAAGAGTCCCTATTTGGGACTTTATCATGATCCAGCTGAGGGTGGAGGCCGATGT 578
Qy 842 lLeuTyrGlyAspSerAsnGlyLeuAspAspSerHisArgGlnLysAspCysPheTrpLe 862
Db 579 ACTGTATGGGACTCCCAATTCCTGGATGACAGTCACCGACAGAAAGACTGCTTTGGCT 638
Qy 862 uLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProProSerLeuSerHisSe 882
Db 639 TCTGGATGCCCTCTCCAGTACACATGATATGGGTGACACCGCTAGACCTCAGTCACTC 698
Qy 882 r-GlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGluG 902

Db 699 TTGGGAACCGCCAGCGCCCTCCACGTGGAGCAGGCTCAGTCCAGAGAGGATGAAG 758
Qy 902 lyAsnHisLeuHisArgTyrSerLysValLeuGluAla 914
Db 759 GAAACCATCTCTCATCGGTACTCTCAAGTCTTGTGAAGGCC 796
RESULT 18
BI414585
LOCUS 60298617ZF1 NCI_CGAP_Lu33 Mus musculus cDNA clone IMAGE:5142401 5',
DEFINITION mRNA sequence.
ACCESSION BI414585
VERSION BI414585.1 GI:15175508
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 929)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM1350 row: p column: 18
High quality sequence start: 29
High quality sequence stop: 840.
Location/Qualifiers
1. 929
/organism="Mus musculus"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:5142401"
/clone_lib="NCI_CGAP_Lu33"
/tissue_type="pooled lung tumors"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was prepared from mRNA obtained from pooled
lung tumors with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAAGTGGAGCGCGCTCTGTGTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 216 a 248 c 260 g 205 t
ORIGIN
Alignment Scores:
Pred. No.: 4.78e-120 Length: 929
Score: 1295.50 Matches: 268
Percent Similarity: 90.42% Conservative: 15
Best Local Similarity: 85.62% Mismatches: 25
Query Match: 23.06% Indels: 8
DB: 13 Gaps: 1
US-09-830-837-6 (1-1052) x BI414585 (1-929)
Qy 371 SerSerArgGlyMetThrTrpGluLeuProGlyTyrGlyArgMetLysProAsp 390
Db 2 TCITCCAGGAGCTGACCTAGTACGAATACCAGAGGCTATGGTCTGTGAAGCCCTGAC 61
Qy 391 lIleValThrTyrGlyAlaGlyValArgGlySerGlyValLysGlyCysArgAlaLeu 410

Db	62	ATGTCACCTATGCTGGAGTGGGGTTCCGGTGTGAAAGGGGCTGCCGTGCACTC	121
QY	411	SerGlyThrSerValAlaSerProValValAlaGlyAlaValThrLeuLeuValSerThr	430
Db	122	TCAGGAACAGTGTGCTCCCTCCAGTGGTGGGCGCTCACCTGTTAGTAAGCACA	181
QY	431	ValGlnLysArgGluLeuValAsnProAlaSerMetLysGlnAlaLeuLeuAlaSerAla	450
Db	182	GTACAGAGCGGAGCTGGTGAATCCTGCCAGTGTGAACAGAGCTTGTAGAGCTCAGCC	241
QY	451	ArgArgLeuProGlyValAsnMetPheGluGlyHisGlyLysLeuAspLeuArg	470
Db	242	CGAGACTTCTGGGTGCACATGTCGACAGAGTCATGGCAAGTGTGATCTGTCGCA	301
QY	471	AlaThrGlnIleLeuAsnSerThrLysProGlnAlaSerLysProSerThrIleAsp	490
Db	302	GCTTATACAGTCTCCAGAGCTATAAACCGGAGGAGCTGAGTCTGCTACATCGAC	361
QY	491	LeuThrGluCysProThrMetThrProThrCysSerGlnProIleThrThrGlyMet	510
Db	362	CTGACTGAGTGTCCCTACATGGCCCTACTGCTCCAGCCCTACTACTATGAGGAATG	421
QY	511	ProThrValAlaValThrIleLeuAsnGlyMetGlyValThrGlyArgIleValAsp	530
Db	422	CCAAACAATCGTTAATGTCAACATCTCAATGGCATGGCGTCCACAGGAAGAAT-GTGGAT	480
QY	531	LysProAspThrGlnProThrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSer	550
Db	481	AAGCCTGAGTGGCAGCCCTATTTACCACAGAATGGAGACACATTTGAAGTGGCTTCTCC	540
QY	551	TyrSerSerValLeuThrProThrSerGlyThrLeuAlaIleSerIleSerValThrLys	570
Db	541	TACTCTCAGTGTGTGGCCCTGGTTCAGTTTACCTTGCCTCCATCTCTGTGACCAAG	600
QY	571	LysAlaAlaSerThrGluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerPro	590
Db	601	AAGGCAGCTTCTGGGAGGAGCATGCTCAGGGCCACATCATCATCAGTGGGCTC-CCA	659
QY	591	AlaGluThrGluSer-LysAsnGlyValAlaGluGlnThrSerThrValLysLeuProIle	610
Db	660	GCAGACAGAGTTACACAGTGTGGCGGACACATTCACCGTAGAGTGGCCATCA	719
QY	610	sValLysIleIleProThrProThrArgSerLysArgValLeuThrAsp-GlnThrHisA	630
Db	720	GGTGAAGATCATTTCCAC-CCTCCTCGGAGCAGAGAGTCTCTGGGACCATGTACCA	778
QY	630	snLeuArgTyrProProGlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProL	650
Db	779	AC---TCCGTACGACCTGCTACTTCCCGAGGACAACTTGGGATGAAGATGACCCCTT	835
QY	650	euAsp-TripAsnGlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHis-LeuAr	669
Db	836	TAGACTTGCATTTGGGAGCAGGTCAAGAGCCAAATTCAGGGAATGTACAGCATCTTGG	895
QY	669	gSerMetGlyTyrPheValGluValLeuGly	679
Db	896	CAGCATGGGGCTACTCGGGAGGTGTCTGGGG	926
RESULT	19		
LOCUS	B0957796		
DEFINITION	AGENCOURT_8772898 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6381912		
ACCESSION	B0957796		
VERSION	B0957796.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 975)		
TITLE	NIH-MGC http://mgi.nci.nih.gov/.		
	National Institutes of Health, Mammalian Gene Collection (MGC)		

JOURNAL COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgapbs-re@mail.nih.gov Tissue Procurement: DCTD/DTP cDNA Library Preparation: Rubin Laboratory cDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL) Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM2572 row: o column: 01 High quality sequence stop: 535. Location/Qualifiers 1..975 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:6381912" /clone_lib="NIH_MGC_40" /tissue_type="carcinoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: Prostate; vector: pCR37; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)." Note: this is a NIH_MGC Library.
BASE COUNT	218 a 280 c 275 g 201 t
ORIGIN	1 others
Alignment Scores:	
Pred. No.:	7.41e-120 Length: 975
Score:	1294.00 Matches: 252
Percent Similarity:	93.77% Conservative: 4
Best Local Similarity:	92.31% Mismatches: 12
Query Match:	23.04% Indels: 5
DB:	14 Gaps: 2
US-09-830-837-6 (1-1052) x BQ957796 (1-975)	
QY	783 HisAspMetTyrTyrAlaSerGlyCysSerIleAlaLysPheProGluAspGlyValVal 802
Db	2 CATGACATGATTTATGCGTTCAGGTGCAGCATCCGGAAGTTTCCAGAGATGGCGTGTG 61
QY	803 IleThrGlnThrPheLysAspGlnGlyLeuGluValLeuLysGlnGluThrAlaValVal 822
Db	62 ATAACACAGACTTTCAAGGACCAAGGATTTGAGGTTTAAAGCAGGAAACACAGCTGTT 121
QY	823 GluAsnValProIleLeuGlyLeuTyrGlnIleProAlaGluGlyGlyArgIleVal 842
Db	122 GAAACAGTCCCATTTTGGGACTTTATCAGATTCCAGCTGAGGGTGGAGCGCGGATGTA 181
QY	843 LeuTyrGlyAspSerAsnGlyLeuAspSerHisArgGlnLysAspCysPheThrLeu 862
Db	182 CTGTATGGGACTCCAATTGCTTGGATGACAGTCCAGCAGACAGAGACTGCTTTTGGCTT 241
QY	863 LeuAspAlaLeuLeuGlnTyrThrSerTyrGlyValThrProSerLeuSerHisSer 882
Db	242 CTGGATGCTCTCTCCAGTACATCGTATGGGTGACACCGCTAGCTCAGTCACTCT 301
QY	883 GlyAsnArgGlnArgProProSerGlyAlaGlySerValThrProGluArgMetGluGly 902
Db	302 GGGACCGCCAGCGCCCTCCAGTGGAGCAGCTCAGTCACTCCAGAGAGGATGGAAGA 361
QY	903 AsnHisLeuHisArgTyrSerLysValLeuGluAlaHisLeuGlyAspProLysProArg 922
Db	362 AACCATCTTCATCGGTACTCCAAGGTTCTGGAGGCCATTTGGGAGACCCAAACCTCGG 421
QY	923 ProLeuProAlaCysProArgLeuSerThrAlaLysProGlnProLeuAsnGluThrAla 942
Db	422 CCTCTACAGCTGTCCACGCTGTCTTGGGGCCAAAGCCACAGCCTTTTAAACGAGACGGC 481

```

QY 943 ProSerAsnLeuTrpLysHisGlnLysLeuLeuSerIleAspLeuAspLysValValLeu 962
DB 482 CCCAGTAACCTTTGGAAACATCAGAGCTACTCTCCATTGACCTGGACAGGTGGTGA 541
QY 963 ProAsnPheArgSerAsnArgProGlnValArgProLeuSerProGlyGluSerGlyAla 982
DB 542 CCCAACTTCGATCGAATCCCTCAAGTGAGGCCCTTGCTCCCTGGAGAGCGCGCC 601
QY 983 TrpAspIleProGlyGlyIleMetPro-GlyArgTyrAsnGlnGluValGlyGlnThrIl 1002
DB 602 TGGGACATCCYGGAGGATCATGCTGGCCCTACACAGGAGGTGGGCCAGACCAT 661
QY 1002 eProValPheAlaPheLeuGlyAlaMet-ValValLeuAlaPheValValGlnIleA 1022
DB 662 TCCTGGCTTTGCTTCCTGGAGGCATGGGTGCTCTGCTCTCTTTGTTGGGACAAATCC 721
QY 1022 snLysAlaLysSerArgPro-LysArgArgLys---ProArgValLysArgProGlnLeu 1040
DB 722 ACAAGGCGAAGACGAAGCAAGCGGAGGAAGCCCGGGGTGACGCCCGCGCGCAGCT 781
QY 1041 --MetGlnGlnValHisProProLysThrPro 1050
DB 782 CACTGGCCCAAGTTTCACCGCGCCCAAAACCCC 814

RESULT 20
BQ928302
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source

```

859 bp mRNA linear EST 20-AUG-2002
 AGENCOURT_8932924 NCI_CGAP_Mam2 Mus musculus CDNA clone
 IMAGE:6488905 5', mRNA sequence.

BQ928302
 BQ928302.1 GI:22343333
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnlni.gov
 Plate: LLAM14036 row: e column: 02
 High quality sequence stop: 621.
 Location/Qualifiers
 1..859
 /organism="Mus musculus"
 /strain="FVB/N-3"
 /db_xref="taxon:10090"
 /clone="IMAGE:6488905"
 /clone_lib="NCI_CGAP_Mam2"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1: SalI;
 Site:2: NotI; Cloned unidirectionally. Primer: Oligo dr.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"

BASE COUNT 187 a 227 c 255 g 190 t
 ORIGIN

Alignment Scores:
 Pred. No.: 6.6e-120 Length: 859
 Score: 1293.50 Matches: 250
 Percent Similarity: 92.55% Conservative: 11
 Best Local Similarity: 88.65% Mismatches: 19

```

Query Match: 23.03% Indels: 2
DB: 14 Gaps: 1
US-09-830-837-6 (1-1052) x BQ928302 (1-859)
QY 343 TyrGlyThrLeuAsnAsnProAlaAspGlnMetAspValIleGlyValGlyIleAsp 362
DB 3 TATGGCACCTCTGAATAACCTCTGATCAGATGGATGTGATTGGAGTGGGTGCATTGAC 62
QY 363 PheGluAspAsnIleAlaArgPheSerSerArgGlyMetThrThrProGluLeuProGly 382
DB 63 TTGAAGATAACATCCCTCGCTTTCTCCAGGGGAATGACTACCTCCGGAATTTACCAGGA 122
QY 383 GlyTyrGlyArgMetLysProAspIleValThrTyrGlyAlaGlyValArgGlySerGly 402
DB 123 GGTATGGTCTGTGTGAAGCTTGACATGTACCTATGGTCTGGAGTGGGGGTTCGGGT 182
QY 403 ValLysGlyGlyCysArgAlaLeuSerGlyThrSerValAlaSerProValValAlaGly 422
DB 183 GTAAAGGGGGCTGCCGTGCACCTCTCAGGACACAGTGTCTCCAGTGGTGGTGGG 242
QY 423 AlaValThrLeuLeuValSerThrValGlnLysArgGluLeuValAsnProAlaSerMet 442
DB 243 GCCGTACCTTGTGTAGACACAGTACAGAACCGGAGCTGGTGAATCTTCCAGTGTG 302
QY 443 LysGlnAlaLeuIleAlaSerAlaArgLeuProGlyValAsnMetPheGluGlnGly 462
DB 303 AGCAAGCTTTGTATAGCTCAGCCCGAGACATCTCTGGGTCAACATGTTCAGACAAAGT 362
QY 463 HisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrLysProGlnAla 482
DB 363 CATGGCAAGTTGGATCTGCTGCGAGCTTATCAGATCTCAGCAGCTATAAACCGCAGCA 422
QY 483 SerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTrpProTyrCysSer 502
DB 423 AGCCTCAGTCCCTAGCTACATCGACCTGAGTGTCCCTACATGTGGCCCTACTGCTCC 482
QY 503 GlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsnGlyMet 522
DB 483 CAGCCTCTACTATGAGGAATGCCAACCAATGTTAATGTACCATCTCTCAATGGCATG 542
QY 523 GlyValThrGlyArgIleValAspLysProAspTrpGlnProTyrIleuProGlnAsnGly 542
DB 543 GCGCTCACAGGAAGAATTGTGGATAAGCCTGAGTGGCGGACCCCTATTATCCACAGAATGA 602
QY 543 AspAsnIleGluValAlaPheSerTyrSerSerValLeuTrpProTrpSerGlyTyrLeu 562
DB 603 GACAACTTGAAGTGGCCCTCTCTACTCTCAGTGTGTGGCCCTGGTCAAGTTACCTT 662
QY 563 AlaIleSerIleSerValThrLysLysAlaAlaSerTrpGluGlyIleAlaGlnGlyHis 582
DB 663 GCCATCTCCATTTCTGGGACCAAGAGGAGCTTCTCTGGGAAGGCATCGCTCCGGGCCAC 722
QY 583 ValMetIle-ThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGluGln-- 601
DB 723 ATCATGATCCCCGGGGGGTCCCCCAGACAGACAGATTACCCGGGGGGGGCGGAGCA 782
QY 602 -ThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSerLys 621
DB 783 CTGTGACCCCGGGGAAGTTCCTCATCAGGGGGAAGATAAATTCCTCCGCCCTCCGCCGGAG 842
QY 621 sArg 622
DB 843 GCAA 846

RESULT 21
BQ928302
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS

```

736 bp mRNA linear EST 07-MAY-2001
 BG697188
 602660420F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4803509 5',
 mRNA sequence.
 BG697188
 BG697188.1 GI:13963119
 EST.
 EST.

[illegible]

(BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 176 a 173 c 206 g 174 t 4 others
ORIGIN

Alignment Scores:

Pred. No.: 1,41e-119 Length: 733
Score: 1289.00 Matches: 234
Percent Similarity: 97.95% Conservative: 5
Best Local Similarity: 95.90% Mismatches: 5
Query Match: 22.95% Indels: 0
DB: 14 Gaps: 0

US-09-830-837-6 (1-1052) x BM949214 (1-733)

QY 632 ArgTyrProGluGlyThrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAsp 651
Db 1 CGCTACCCACCTGGGTACTTCCCGAGGACAACTTGGGATGAAGAATGACCTTTAGAC 60
QY 652 TrpAsnGlyAspHisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMet 671
Db 61 TGAATGGGACGACGTCACACCACTTCAGGACATGTACAGCATCTGCGACATG 120
QY 672 GlyTyrPheValGluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGly 691
Db 121 GGCTACTTCTGGAGGTGCTCGCGCCCATTCACATGTTTGACGCCACACAGATGCG 180
QY 692 ThrLeuLeuMetValAspSerGluGluGluTyrPheProGluGluIleAlaLysLeuArg 711
Db 181 ACTTTGCTGCTGGTGACAGTGAAGAAGTAGTACTTCCCGAGGAGATTCTTAAGCTGAG 240
QY 712 ArgAspValAsnGlyLeuSerLeuValIlePheSerAspTrpTyrAsnThrSerVal 731
Db 241 AGGGATGTGGCAATGGCCCTTCCCTCGTCATCTTCAGTGACTGGTACACACTTCTGT 300
QY 732 MetArgLysValLysPheTyrAspGluAsnThrArgGlnTrpTrpMetProAspThrGly 751
Db 301 ATGAGAAAAGTGAAGTTTATGATGAAAACACAGGAGTGGTGGATGCCAGACCCGGA 360
QY 752 GlyAlaAsnIleProAlaLeuAsnGluLeuLeuSerValTrpAsnMetGlyPheSerAsp 771
Db 361 GGAGCGAATCCCACTGTAATGAGCTGCTGTGTGTGGAACATGGGGTTCAGTGAC 420
QY 772 GlyLeuTyrGluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCys 791
Db 421 GGCCATATGAGGGGAGTTTCTCTGGCANACCATGACATGACTATCGCTGGGGTGC 480
QY 792 SerIleAlaLysPheProGluAspGlyValValIleThrGlnThrPheLysAspGlnGly 811
Db 481 AGCATCGCCCAAGTTTCCAGAAGATGGCGTGTGATCACACAGACTTTCANGGACCAANGA 540
QY 812 LeuGluValLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyr 831
Db 541 TTGGAGGTCTTAAACAAGACAGACAGTGTGGAAAATGTTCCTATTTGGGGCTTTAT 600
QY 832 GlnIleProAlaGluGlyGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAsp 851
Db 601 CAGATCCATCTGAAGTGGAGCGCGGATCGTGTGTATGGAGACTCCCAACTGCTTGGAT 660
QY 852 AspSerHisArgGlnLysAspCysPheThrPheLeuLeuAspAlaLeuLeuGlnTyrThrSer 871
Db 661 GACAGTCACAGACAGAGGACTGCTTTTGGCTTCTGGATCGCTCTTTCAGTACACATCC 720
QY 872 TyrGlyValThr 875
Db 721 TATGGCGTGACC 732

RESULT 23

AUI39053

LOCUS

DEFINITION AUI39053 809 bp mRNA linear EST 02-AUG-2002
AUI39053 PLACE1 Homo sapiens cDNA clone PLACE1009836 5', mRNA
sequence.

ACCESSION

AUI39053 GI:11000574

VERSION

EST.

KEYWORDS

human.

SOURCE

Homo sapiens

ORGANISM

REFERENCE 1 (bases 1 to 809)
AUTHORS Ota.T., Nishikawa.T., Suzuki.Y., Ishii.S., Saito.K., Kawai.Y., Yamamoto.J., Wakamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and Isogai.T.

TITLE

HRI human cDNA project

JOURNAL

Unpublished (2000)

COMMENT

Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp

HRI human cDNA project: 5' - & 3' - end one pass sequencing: Helix Research Institute; cDNA library construction: Department of Virology, Institute of Medical Science, University of Tokyo, and Helix Research Institute.

FEATURES

source

1. 809
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="PLACE1"
/clone_lib="PLACE1"
/tissue_type="placenta"
/note="Vector: pME18SFL3"

BASE COUNT 211 a 194 c 213 g 185 t 6 others
ORIGIN

Alignment Scores:

Pred. No.: 1.55e-118 Length: 809
Score: 1279.50 Matches: 248
Percent Similarity: 93.28% Conservative: 2
Best Local Similarity: 92.54% Mismatches: 12
Query Match: 22.78% Indels: 6
DB: 9 Gaps: 1

US-09-830-837-6 (1-1052) x AUI39053 (1-809)

QY 536 ProTyrLeuProGlnAsnGlyAspAsnIleGluValAlaPheSerTyrSerSerValLeu 555
Db 3 CCTATTTGCCACAGAACGAGACAACTTGAAGTTGGCTTCTCTACTCTCGGTCTTA 62
QY 556 TrpProTyrSerGlyTyrLeuAlaIleSerIleSerValThrLysLysAlaLaSerTyr 575
Db 63 TGGCCTTGGTGGGCTACCTGGCCATCTCCATTTCTGTGCCAAGAAAGCGGCTTCTGG 122
QY 576 GluGlyIleAlaGlnGlyHisValMetIleThrValAlaSerProAlaGluThrGluSer 595
Db 123 GAAGGCATTGCTCAGGGCCATGTCATGATCATCTGGTGTCCCGACAGACAGAGTCA 182
QY 596 LysAsnGlyAlaGluGlnThrSerThrValLysLeuProIleLysValLysIleIlePro 615
Db 183 AAAAAATGGTGCAGAACAGACTTCAACAGATAAGCTCCCATTAAGGTGAAGATAATTCCT 242
QY 616 ThrProArgSerLysArgValLeuTyrAspGlnTyrHisAsnLeuArgTyrProPro 635
Db 243 ACTCCCCCGGAGCAAGAGAGTTCCTGGGATCAGTACACAACTCCGCTATCCACCT 302
QY 636 GlyTyrPheProArgAspAsnLeuArgMetLysAsnAspProLeuAspTrpAsnGlyAsp 655
Db 303 GCGTATTTCCCGAGGATAATTTAAGGATGAAGATGACCTTTAGACTTGAATGCTGAT 362
QY 656 HisIleHisThrAsnPheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheVal 675
Db 363 CACATCCACCAATTTCCAGGATATGTACCAGCATCTGAGAAGCATCTGGCTACTTTGTA 422

QY 676 GluValLeuGlyAlaProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMet 695
 Db 423 GAGGTCCTCGGGCCCCCTTCACGGTTTTCATGCCAGTCAGTATGGCACTTTGCTGATG 482
 QY 696 ValAspSerGluGluTyrPheProGluGluIleAlaLysLeuAraGaspValAsp 715
 Db 483 GTGGACATGGGGAGGAGTACTTCCTCAAGAGATGCCAAGCTCCGGAGGGAGCTGGAC 542
 QY 716 AsnGlyLeuSerLeuValIlePheSerAspTyrPheAsnThrSerValMetArgLysVal 735
 Db 543 AACGNCCTCGCTCGCTCATCTTCAGTCACTGGTACACACTTCTGTATGACAAAGTG 602
 QY 736 LysPheTyrAspGluAsnThrArgGlnTyrPheProAspThrGly--GlyAlaAsnI 755
 Db 603 AAGTTTATGATGAACAACAAGCAGTGGTGGATCCCGATACCGGGAGGAGCTAACCA 662
 QY 755 leProAlaLeuAsnGluLeuLeuSerValTyrPheAsnMetGlyPheSerAspGlyLeuTyrG 775
 Db 663 TCCAGCTCTGAATGAANTGCTGCTGTGTGGNACATGGGGGTCAACGATGGCTGTATG 722
 QY 775 lu-GlyGluPheThrLeuAlaAsnHis-AspMetTyrTyrAlaSer-GlyCysSerIleA 794
 Db 723 AAAGGGGAGTCAACCTGGGCAACCATGGACATGATATTANGGCTCAAGGTGCAACCAAC 782
 QY 794 laIys---PhePro 797
 Db 783 CCCGAANTTTCCA 796

RESULT 24
 BI661522
 LOCUS
 DEFINITION 724 bp mRNA linear EST 12-SEP-2001
 603305833F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:5351337 5',
 mRNA sequence.
 BI661522
 BI661522.1 GI:15575758
 EST.
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 724)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)
 Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
 Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lni.gov>
 Plate: L14M11893 row: b column: 10
 High quality sequence stop: 723.
 Location/Qualifiers
 1..724
 /organism="Mus musculus"
 /strain="NMRI"
 /db_xref="taxon:10090"
 /clone="IMAGE:5351337"
 /clone_lib="NCI_CGAP_Mam4"
 /tissue_type="tumor, gross tissue"
 /dev_stage="5 months"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-SPORT6; Site:1; SalI;
 Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT.
 Library constructed by Life Technologies. Investigators
 providing samples: Lothar Hennighausen/Priscilla Furth.
 NIH Reference for transgenic model: Li et al., Cell Growth
 and Differentiation 7, 3-11 (1996)."

BASE COUNT 175 a 212 c 181 g 156 t

ORIGIN

Alignment Scores:
 Pred. No.: 4.84e-116 Length: 724
 Score: 1254.00 Matches: 228
 Percent Similarity: 98.34% Conservatives: 9
 Best Local Similarity: 94.61% Mismatches: 3
 Query Match: 22.33% Indels: 1
 DB: 13 Gaps: 0

US-09-830-837-6 (1-1052) x BI661522 (1-724)

QY 461 GlnGlyHisGlyLysLeuAspLeuLeuArgAlaTyrGlnIleLeuAsnSerTyrIysPro 480
 Db 2 CAAGTCTATGCAAGTTGGATCTGTGCGAGCTTATCAGATCTCTCAGCAGCTATAAACC 61
 QY 481 GlnAlaSerLeuSerProSerTyrIleAspLeuThrGluCysProTyrMetTyrProTyr 500
 Db 62 CAGGCAAGCCTGAGTCTAGCTACATCGACCTGACTGAGTGTCCCTACATGTGGCCCTAC 121
 QY 501 CysSerGlnProIleTyrTyrGlyMetProThrValValAsnValThrIleLeuAsn 520
 Db 122 TGCTCCCAAGCTTACTACTATGGAGGATGCCACATCGTTAATGTCAACCTCTCAAT 181
 QY 521 GlyMetGlyValThrGlyArgIleValAspLysProAspTyrProGlnProTyrLeuProGln 540
 Db 182 GGCATGGCGGTACAGAGAGATTTGGATAGCTCTGAGTGGCGACCTATTATTACACAG 241
 QY 541 AsnGlyAspAsnIleGluValAlaPheSerTyrSerValLeuTyrProTyrPserGly 560
 Db 242 AATGGAGACACATTTGAAGTGGCTTCTCTACTCTCTCAGTGTGTGGCCCTGTGTCAGGT 301
 QY 561 TyrLeuAlaIleSerIleSerValThrLysLysAlaAlaSerTyrGluGlyIleAlaGln 580
 Db 302 TACCTTGGCATCTCCATTTCTGTGACCAAGAGGACAGCTCTCTGGAGGACATCGCTCAG 361
 QY 581 GlyHisValMetIleThrValAlaSerProAlaGluThrGluSerLysAsnGlyAlaGlu 600
 Db 362 GGCCACATCATGATCAGTGGCGTCCCAGCAGACAGAGTATACACAGTGTGGTGGAG 421
 QY 601 GlnThrSerThrValLysLeuProIleLysValLysIleIleProThrProArgSer 620
 Db 422 CACACTTCCACCGTGAAGCTGCCATCAAGTGAAGATCATTCACACCTCTCTCGGAGC 481
 QY 621 LysArgValLeuTyrPaspGlnTyrHisAsnLeuArgTyrProGlyTyrPheProArg 640
 Db 482 AAGAGAGTCTCTGGGACCATACCAACCTCCGCTACCCACCTGGTCTCTCCCGAG 541
 QY 641 AspaAsnLeuArgMetLysAsnAspProLeuAspTyrPaspGlyAspHisIleThrAsn 660
 Db 542 GACACTTGGCGATGAGATGATGACCTTTAGACTGGATGGCAGCAGCAGTCCACACCAAC 601
 QY 661 PheArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPheValGluValLeuGlyAla 680
 Db 602 TTCAGGACATGTACCAAGCATCTGGCAGCATGGGCTACTCTGTGGAGGTGCTGGCGCC 661
 QY 681 -ProPheThrCysPheAspAlaSerGlnTyrGlyThrLeuLeuMetValAspSerGluG 700
 Db 662 CCCATTACATGTTTGGACGCCACACAGTATGGCACTTTGCTGTGGTGGACAGTGAGGA 721
 QY 700 u 700
 Db 722 A 722

RESULT 25
 LOCUS
 BI091766
 DEFINITION 880 bp mRNA linear EST 20-JUN-2001
 602858816F1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:5000283 5',
 mRNA sequence.
 BI091766
 BI091766.1 GI:14510096
 EST.
 KEYWORDS
 SOURCE
 human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 880)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: Incyte Genomics, Inc.

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLAM1031 row: o column: 04

High quality sequence stop: 726.

FEATURES

source

1..880
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5000283"
/cell_line="NIH-MGC_10"
/cell_line="MGC36"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.5 kb. Library prepared by Life
Technologies."

BASE COUNT 214 a 236 c 239 g 191 t

ORIGIN

Alignment Scores:

Pred. No.: 2,81e-115 Length: 880
Score: 1248.00 Matches: 258
Percent Similarity: 89.86% Conservative: 8
Best Local Similarity: 87.16% Mismatches: 22
Query Match: 22.22% Indels: 10
DB: 13 Gaps: 3

US-09-830-837-6 (1-1052) x BI091766 (1-880)

QY 655 AsphHisIleHisThrAsnPhaArgAspMetTyrGlnHisLeuArgSerMetGlyTyrPhe 674
DB 3 GATCATCCACACCAATTCAGGATATGTACAGCATCTGAGAGCATGGGCTACTTT 62
QY 675 valgluvalleuglyalaprophetrcyspheaspalaserglnthyrlthleu 694
DB 63 GTAGAGTCTCGGGGCGCCCTTCAGTGTGTTGATGCCAGTCAGTATGCCACTTTGCTG 122
QY 695 MetValAspSerGluGluTyrPheProGluGluIleAlaLysLeuArgAspVal 714
DB 123 ATGGTGACAGTGAGGAGGAGTACTTCCCTCGAGAGATGCCAAGCTCGGAGGACGTG 182
QY 715 AspAsnGlyLeuSerLeuValIlePheSerAspTyrTrpAsnThrSerValMetArgLys 734
DB 183 GACACGGCTCTCGCTGCTCATCTTCAGTGACTGTGTACACACTTCTGTATGAGAAA 242
QY 735 vallyspheTyrAspGluAsnThrArgGlnTyrTrpMetProAspThrGlyGlyAlaAsn 754
DB 243 GTGAAGTTTATGATGAAAAACACAGGCAGTGTGGATGCCGATACCCGAGGAGCTAAC 302
QY 755 IleProAlaLeuAsnGluLeuSerValTrpAsnMetClyPheSerAspGlyLeuTyr 774
DB 303 ATCCAGCTCTGAATGAGTGTCTGTGTGGAACATGGGGTTACCGCATGGCTGTAT 362
QY 775 GluGlyGluPheThrLeuAlaAsnHisAspMetTyrTyrAlaSerGlyCysSerIleAla 794
DB 363 GAAGGGAGTTCACCTCGCCACCATGACATGATATGCGTCAGGGTCAGCATCGCG 422
QY 795 LysPheProGluAspGlyValIleThrGlnThrPheLysAspGlnGlyLeu-GluVa 814
|||||

Db

423 AAGTTTCCAGAGATGGCGCTGATAACACAGACTTTCAGGACCAAGGATGGCAGGT 482
QY 814 lLeuLysGlnGluThrAlaValValGluAsnValProIleLeuGlyLeuTyrGlnIlePr 834
DB 483 TTTAAAGCAGAAACAGCAGTGTGTAACACGTCCTCATTTTGGGACTTTATCAGATTCC 542
QY 834 oAlaGluGlyGlyGlyArgIleValLeuTyrGlyAspSerAsnCysLeuAspAspSerHi 854
DB 543 AGCTGAGGGTGGAGCGCGGATTGCTGTATGGGACTCCCAATTGCTTGGATGACAGTCA 602
QY 854 sArgGlnLysAspCysPheTrpLeuLeuAspAlaLeuLeuGlnTyrThrSerTyrGlyVa 874
DB 603 CCACAGAGAGGACTGCTT-TGGCTTCTGGATGCCCTCTCCAGTACACATCGTATGGGT 661
QY 874 lThrProProSerLeuSerHisSerGlyAsnArgGlnArgProProSerGlyAlaGlySe 894
DB 662 GACACCGCTTAGCCTCAGTCACCTCTGGGAACCG-CAGCGCTCCCTAGTGAAGCAGCTC 720
QY 894 rValThrPro---GluArgMetGluGlyAsnHisLeu---HisArgTyrSerLysValle 912
DB 721 AGTCAGTCCAGAGAGGAAATCGGAAGGAAACGATCTTCCATCGGACTCCAAGTTCTCT 780
QY 912 u-GluAlaHisLeuGly---AspProLysProArg-ProLeuProAla-CysProArg-L 930
DB 781 CGCAGGCCCCAATTCGCCAAGAACCCACAACTCGCGCTCTAGCAACCCCTGCCACGCC 840
QY 930 euSerTrpAlaLysProGlnProLeuAsnGluThrAla 942
DB 841 TGCCATGGCGCCAGCACCACGCTCTACACCCACACCGCC 878

Search completed: June 2, 2003, 03:52:21

Job time : 3431 secs

THIS PAGE BLANK (USPTO)